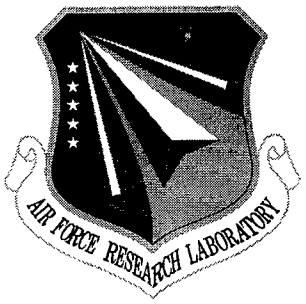
AFRL-IF-RS-TR-2001-76 Vol III (of four)
Final Technical Report
May 2001



EXPLORING A THEORY DESCRIBING THE PHYSICS OF INFORMATION SYSTEMS, CHARACTERIZING THE PHENOMENA OF COMPLEX INFORMATION SYSTEMS

Zetetix

Sponsored by
Defense Advanced Research Projects Agency
DARPA Order No. K177

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

The views and conclusions contained in this document are those of the authors and should not be interpreted as necessarily representing the official policies, either expressed or implied, of the Defense Advanced Research Projects Agency or the U.S. Government.

20010713 056

AIR FORCE RESEARCH LABORATORY
INFORMATION DIRECTORATE
ROME RESEARCH SITE
ROME, NEW YORK

This report has been reviewed by the Air Force Research Laboratory, Information Directorate, Public Affairs Office (IFOIPA) and is releasable to the National Technical Information Service (NTIS). At NTIS it will be releasable to the general public, including foreign nations.

AFRL-IF-RS-TR-2001-76 Vol. III (of four) has been reviewed and is approved for publication.

APPROVED: Deborah A. Cerino

DEBORAH A. CERINO
Project Engineer

FOR THE DIRECTOR: James A. Collins

JAMES A. COLLINS, Acting Chief
Information Technology Division
Information Directorate

If your address has changed or if you wish to be removed from the Air Force Research Laboratory Rome Research Site mailing list, or if the addressee is no longer employed by your organization, please notify AFRL/IFTD, 525 Brooks Road, Rome, NY 13441-4505. This will assist us in maintaining a current mailing list.

Do not return copies of this report unless contractual obligations or notices on a specific document require that it be returned.

EXPLORING A THEORY DESCRIBING THE PHYSICS OF INFORMATION SYSTEMS, CHARACTERIZING THE PHENOMENA OF COMPLEX INFORMATION SYSTEMS

Scott Young Harmon

Contractor: Zetetix
Contract Number: F30602-00-C-0104
Effective Date of Contract: 06 April 2000
Contract Expiration Date: 05 October 2000
Short Title of Work: Exploring a Theory Describing the Physics of Information Systems, Characterizing the Phenomena of Complex Information Systems
Period of Work Covered: Apr 00 - Oct 00

Principal Investigator: Scott Young Harmon
Phone: (818) 991-0480
AFRL Project Engineer: Deborah A. Cerino
Phone: (315) 330-1445

APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

This research was supported by the Defense Advanced Research Projects Agency of the Department of Defense and was monitored by Deborah A. Cerino, AFRL/IFTD, 525 Brooks Road, Rome, NY.

REPORT DOCUMENTATION PAGE	Form Approved OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188), Washington, DC 20503.

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE	3. REPORT TYPE AND DATES COVERED
	MAY 2001	Final Apr 00 - Oct 00

4. TITLE AND SUBTITLE
EXPLORING A THEORY DESCRIBING THE PHYSICS OF INFORMATION SYSTEMS, CHARACTERIZING THE PHENOMENA OF COMPLEX INFORMATION SYSTEMS

5. FUNDING NUMBERS
C - F30602-00-C-0104
PE - 63760E
PR - IAST
TA - 00
WU - P1

6. AUTHOR(S)
Scott Young Harmon

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
Zetetix
PO Box 2640
Agoura CA 91376-2640

8. PERFORMING ORGANIZATION REPORT NUMBER
N/A

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
Defense Advanced Research Projects Agency
3701 North Fairfax Drive
Arlington VA 22203-1719

Air Force Research Laboratory/IFTD
525 Brooks Road
Rome New York 13441-4505

10. SPONSORING/MONITORING AGENCY REPORT NUMBER
AFRL-IF-RS-TR-2001-76
Vol III (of four)

11. SUPPLEMENTARY NOTES
Air Force Research Laboratory Project Engineer: Deborah A. Cerino/IFTD/(315) 330-1445

12a. DISTRIBUTION AVAILABILITY STATEMENT
APPROVED FOR PUBLIC RELEASE; DISTRIBUTION UNLIMITED.

12b. DISTRIBUTION CODE

13. ABSTRACT *(Maximum 200 words)*

This project accomplished all of its objectives: document a theory of information physics, conduct a workshop on planning experiments to test this theory, and design experiments that validate this theory. Information physics proposes quantitative relationships between observable information flows and changes in the content information systems maintain. This theory explains all flows within information systems as either diffusive or force-driven. The forces driving information flows arise from the existence of goal content. The workshop participants discussed various theories and considered experiments that characterize the macroscopic phenomena underlying complex information system behavior. These participants identified experimental opportunities that exploit existing databases, execute simulations and conduct traditional controlled experiments. They recommended that focussed experiments to test theories explaining information system phenomena were feasible today. The experiment plan builds upon the workshop's result and proposes experiments that measure information device energy dissipation, test the independence of symbol execution work from device efficiency, measure information diffusion rates in information systems, and measure force-driven information flows. These experiments are both technically and programmatically feasible. When validated, the proposed theory can guide designers to reliably build more effective, secure and predictable information systems.

14. SUBJECT TERMS	15. NUMBER OF PAGES
Information Physics, Physics of Computation, Information Theory, Thermodynamics	80
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UL

TABLE OF CONTENTS

Abstract 1

Introduction to the Workshop 2

Workshop Agenda 4

Introduction to Workshop Participants 5

Workshop Presentations 10

Keynote Presentation: All Roads Lead to Rome 10

Information Assurance Mathematics 11

Nature of Interactions and Emergent Properties 13

Categorical Modeling of Information Assurance 13

CyberLogic: A Logical Characterization of Macroscopic Phenomena of Complex Information Systems 15

Is Fisher Information a Governing Concept in Cyberspace? 17

Kolmogorov Complexity as a Conserved Parameter of Information Assurance 19

A Physical Model of the Behavior of Information Systems 20

Synopsis of Current Directions in Cartography of Cyberspace 25

Information Assurance – A Complex Systems Perspective 26

Working Session Results 28

Theoretical Description of Information System Phenomena 28

Phenomena of Information Systems 29

Existing Theory Base Describing Information Systems 35

Theory Relevant to IA 36

Outstanding IA Theory Challenges 41

Experimental Validation of Information System Theories 45

Existing Data Sources 46

Simulation Resources 51

Experimental Facilities 53

Experimental Challenges 55

Experimental Guidance 58

Validation of IA-Related Theories 62

UML as Theory Notation 64

Conclusions and Recommendations 65

Workshop Conclusions 65

Workshop Recommendations 66

References 68

Acknowledgements 68

LIST OF TABLES

Table 1.	Properties Describing Information System Structure.	29
Table 2.	Properties Describing Information System Dynamics.	30
Table 3.	Properties Describing Information Characteristics.	31
Table 4.	Possible Relationships that Exist between Information System Properties.	33
Table 5.	Contributions of Information System-Related Theory Base to IA Problems.	37
Table 6.	Correspondence between Individual Participant Theories and Major Theoretical Bases.	40
Table 7.	Summary of Outstanding IA Challenges.	42
Table 8.	Common Capabilities of the Support Infrastructure for the IA Laboratory.	54
Table 9.	Combined Computing Capabilities Housed by the IA Laboratory.	54
Table 10.	Possible Approaches to Validate Individual Participant Theories.	62

ABSTRACT

This workshop focused upon revitalizing research in complex information system physics by assessing the current state of understanding, identifying opportunities for theoretical refinement and verification by experimentation, and encouraging communications among researchers. To accomplish these objectives, the workshop sought to

- Provide a forum to encourage communication and collaboration among researchers from disparate disciplines to improve the coherence and coverage of existing theoretical work
- Assess the existing body of empirical evidence on information system behavior to determine its applicability and limitations for testing current and future theory
- Identify opportunities for additional controlled experiments and observations of existing information system behavior, in the context of information assurance (IA), to test present and future theory

This workshop began with the participants presenting their concepts that explain macroscopic information system phenomena. Then, the participants conducted several working sessions examining the current theoretical base and assessing the possibilities for experimental validation of the existing and future theoretical developments in information system physics.

By the workshop's end, its participants concluded unanimously that better knowledge of the scientific fundamentals of information systems will improve the state of information assurance practice and that sufficient theory describing information system phenomena exists in narrow areas so that it could be further explored with small focused experiments. The participants cautioned that the science of information systems is indeed in its infancy. The lack of a common vocabulary and the uneven characterization of the mappings of microscopic variables to macroscopic phenomena indicate this immaturity. However, substantial theoretical affinities exist with well developed disciplines (e.g., mathematics, physics, economics, sociology, evolution). These affinities could be leveraged to rapidly accelerate the maturing of the science describing information system behavior. On the other hand, despite these affinities, none of these well establish disciplines are enough without further development. Building upon these conclusions, the workshop participants formulated several recommendations:

- Aggressively pursue the science explaining information system phenomena
- Emphasize experimentation in this pursuit
- Define consistent terminology to promote communications among researchers
- Build upon existing knowledge to minimize redundancy
- Identify and fill theory holes to broaden its applicability and practicality
- Link scientific discovery to practical development to hasten application
- Prioritize information assurance problems to focus theoretical development

INTRODUCTION TO THE WORKSHOP

Recent proliferation of ever larger and more complex information systems has increased the vulnerability of these systems to chaotic performance, catastrophic failure, and intentional disruption and compromise. A primary reason is that the fundamental behavior of complex information systems appears to be poorly understood. At best, failure to improve this situation will ultimately limit the level of complexity that information systems can reliably and practically achieve. At worst, an ill understood worldwide information system fabric could become unstable, falter without warning, and fail catastrophically. These consequences can result in the impairment of critical services, creation of global economic disruption, and, possibly, loss of life.

In the quest to achieve better fundamental understanding of information systems, considerable research has been done to unify the disciplines of physics and computational theory. This large body of accomplishment provides a rich resource for building a comprehensive and predictive understanding of the macroscopic behavior of complex information systems. The focus of this workshop was to revitalize research in complex information systems by assessing the current state of understanding, identifying opportunities for theoretical refinement and verification by experimentation, and encouraging communications among researchers. The products of this workshop should contribute to building a better understanding of the phenomena underlying the behavior of complex information systems that will ultimately lead to dramatically improved security and reliability.

To accomplish these objectives, the workshop sought to

- Provide a forum to encourage communication and collaboration among researchers from disparate disciplines to improve the coherence and coverage of existing theoretical work
- Assess the existing body of empirical evidence on information system behavior to determine its applicability and limitations for testing current and future theory
- Identify opportunities for additional controlled experiments and observations of existing information system behavior, in the context of information assurance (IA), to test present and future theory

Participation in this workshop was by invitation only. The workshop call for participation solicited the involvement of researchers in such areas as information mechanics, physics of computation, thermodynamics of communications and computation, physics of information, computational mechanics, and IA.

Technical papers, both theoretical and empirical, were solicited that summarized

- The existing theories describing the macroscopic phenomena of complex information systems by (1) identifying what phenomena their theory addresses, (2) defining the characteristics of the information systems they are describing, (3) characterizing the conditions under which

their theories likely best apply, and (4) discussing any empirical support that may exist for their theories

- The empirical data that have already been collected on complex information system behavior by (1) identifying the specifics of the experiments or observations that would define the limits of their validity, (2) characterizing the error sources and magnitudes of likely errors, (3) summarizing the hypotheses that their data confirms or refutes, and (4) discussing any other conclusions that their observations might suggest

To promote communications and interchange among workshop participants, the workshop schedule consisted of one invited presentation, several contributed presentations, and seven working sessions. The emphasis on working sessions was on identifying and characterizing the opportunities to empirically validate present and future complex information system theories.

WORKSHOP AGENDA

August 29 - Morning		
Introduction and Administrative Comments	Nicholson	0830-0835
Welcome and DARPA Objectives	Skroch	0835-0850
Participant Introductions	All	0850-0915
All Roads Lead to Rome	Toffoli	0915-1000
Break		1000-1030
Information Assurance Mathematics	Benzinger	1030-1100
Nature of Interactions and Emergent Properties	Belyavin	1100-1130
Categorical Modeling of Information Assurance	Clarke	1130-1200
Lunch		1200-1330
August 29 - Afternoon		
CyberLogic	Saidi	1330-1400
Is Fisher Information a Governing Concept in Cyberspace?	Cox	1400-1430
Kolmogorov Complexity as a Conserved Parameter of Information Assurance	Evans & Bush	1430-1500
Break		1500-1530
A Physical Model of the Behavior of Information Systems	Harmon	1530-1600
Unscheduled Presentations		1600-1630
Working Session 1: Presentation Discourse	All	1630-1700
August 30 - Morning		
Working Session 2: Coverage of Current Theory	All	0830-1000
Break		1000-1030
Working Session 3: Theory to Experiment	All	1030-1200
Lunch		1200-1330
August 30 - Afternoon		
Working Session 4: Existing Experiment Coverage	All	1330-1500
Break		1500-1530
Working Session 4: Experiment Opportunities	All	1530-1700
August 31 – Morning		
Working Session 6: Experiment Priorities	All	0830-1000
Break		1000-1030
Working Session 7: Workshop Recommendations	All	1030-1130
Concluding Remarks	Harmon	1130-1200
Workshop Adjourns		1200

INTRODUCTION TO WORKSHOP PARTICIPANTS

Prior to any presentations, each participant identified themselves and their affiliation, and very briefly described their interest in the workshop topic. Subsequent to the workshop, each participant provided their biography. These are given below.

Andrew Belyavin

Dr Andrew Belyavin is Principal Scientist in the DERA Centre for Human Sciences (CHS), specialising in statistical analysis and human performance modelling. He graduated with a BA in Mathematics at the University of Cambridge in 1971, and a Ph.D. in Applied Statistics at the University of Reading in 1981. He joined the RAF Institute of Aviation Medicine in 1975 as a Senior Research Fellow studying multivariate methods for small samples. He was appointed head of the Experimental Design and Analysis section in 1978, becoming Head of Mathematical Modelling, Statistics and Computing in 1987. He moved to CHS on its formation in 1994, leading the Biometrics and Ergonomics group, providing consultancy on the statistical problems of surveys and designed experiments, and contributing to the development of models of aspects of human performance, including occupational stress, fatigue, whole-body thermal response, and workload. In 1995, he became project manager for the Integrated Performance Modelling Environment (IPME), a large project funded by the Ministry of Defence to model the overall effectiveness of systems with human operators. He is currently working on micro-models for IPME, including quantitative models of the effects of various stressors upon performance and leads the general modelling capability group at CHS, which includes the development of manpower models and other related modelling topics.

Lee Benzinger

Lee A. Benzinger has worked in the area of high assurance systems and software for nearly 20 years. For the last two years she has been working on DARPA projects related to intrusion detection, intrusion tolerance, and information assurance. In addition, she provided the network and security management architecture for the next generation DoD mobile communications system under Phase I of the Joint Tactical Radio System Program (JTRS). Before coming to NAI Labs in January 1999, she worked at Lockheed Martin Western Development Laboratories (formerly Loral and Ford Aerospace) as a security engineer/researcher and analyst/consultant for large distributed communication systems. These systems include the Globalstar satellite telecommunications system and the Contingency Airborne Reconnaissance System. Her primary research was the development of the WDL theory, a mathematical theory for modeling systems, components, and component and policy composition. She has applied this theory in the analysis of a large communication system with multilevel segments that was undergoing upgrade.

Currently, she is the PI on two DARPA contracts and manages the Santa Clara office of NAI Labs.

Steven Bush

Dr. Stephen F. Bush is a Computer Scientist at General Electric Research and Development in Niskayuna, NY. Dr. Bush conducts research in advanced networking concepts. He has numerous patents pending in the area of network security and vulnerability analysis and provisional patents for active networks in both terrestrial and space based communication systems. He is completing a book to be published by Kluwer/Plenum Academic Publishers in 2001 titled "Active Networks and Active Network Management: A Proactive Management Framework".

Dr. Bush received his BS in Electrical and Computer Engineering from Carnegie Mellon University and MS in Computer Science from Cleveland State University. Before joining at General Electric Research and Development, he was a researcher at the Information and Telecommunications Technologies Center at the University of Kansas where he contributed to the DARPA Rapidly Deployable Radio Networks Project. Dr. Bush completed his Ph.D. research at the University of Kansas where he received a Strobel Scholarship Award. He received the award of Achievement for Professional Initiative and Performance for his work as Technical Project Leader at General Electric Information Systems in the areas of network management and control while pursuing his Ph.D.

Thomas Clarke

Dr. Thomas L. Clarke is a Senior Scientist at the Institute for Simulation and Training at the University of Central Florida. He has more than 20 years research experience involving propagation modeling, digital processing of acoustic signals, statistical analysis, and numerical modeling. He has five patents.

Dr. Clarke has published extensively in professional journals and is affiliated with the Institute of Electrical and Electronics Engineers, the Audio Engineering Society, and the American Physical Society. He received a BS in Mathematics from Florida International University in 1973, an MS in Applied Mathematics from The University of Virginia in 1975, and a Ph.D. in Applied Mathematics from the University of Miami in 1982.

He worked as a mathematician at the Atlantic Oceanographic and Meteorological Laboratory in Miami, FL, from 1975 to 1987. He has been the Florida Principal Mathematician at the Institute for Simulation and Training, University of Central Florida, in Orlando, FL, since June 1988.

Roger Cox

Dr. Roger Cox has spent most of his career as a systems engineer, mathematical modeller and algorithm designer in various engineering disciplines. He received a BA in Mathematics and an MSE in Environmental Engineering from Johns Hopkins University. He then obtained a Ph.D. in Civil Engineering from Cornell University. From 1984 to 1990 Dr. Cox worked at AT&T Bell Laboratories in Holmdel, NJ where he was responsible for systems engineering and network

architecture for transmission networks and private line services. He joined Sandia National Laboratories in Albuquerque, NM in 1990 where he developed architectures and designs for environmental decision support systems and worked on telecommunications network reliability, aviation safety, and miscellaneous scientific and technical projects. Recently, he became interested in extending Roy Frieden's mathematical machinery for deriving physical laws from Fisher information to various application areas.

Scott Evans

Scott C. Evans is an Electrical Engineer at General Electric Research and Development in Niskayuna, NY. He conducts research in advanced communications and networking concepts. He has numerous patents pending in the area of wireless and power line communications and information security.

Mr. Evans received his BS in Electrical Engineering from Virginia Tech and MS in Electrical Engineering from the University of Connecticut. He is currently pursuing a PHD in Electrical Engineering at Rensselaer Polytechnic Institute. Before joining General Electric Research and Development, Mr. Evans was a nuclear-trained Submarine Officer in the United States Navy and a design engineer at GE Industrial Systems.

Michael Frentz

Michael Frentz is a Lead Engineer and Deputy Director of the Information Security Department at the BBN Technologies (one of two research divisions of Verizon). His areas of expertise are in systems development, modeling, software development management, and signal processing. Mr. Frentz received a B.E.E. in Electrical Engineering and an MS in Physics from the Catholic University of America.

Mr. Frentz laid much of the theoretical and experimental scientific groundwork in the low frequency multistatic active (LF-MSA) sonar and was responsible for the first three generations of fielded aircraft-based systems in that area. Mr. Frentz was the software manager for the development of the MISSI Certification Authority Workstation (the PKI component of DMS) and has a patent pending in the market-driven filtering of network communications. He is the PI on the Cartography of Cyberspace effort (which also includes: Ceilyn Boyd, Dave Mankins, Bill Nelson, Wally Feurzeig, and Oliver Selfridge).

Vipin Gopal

Dr. Vipin Gopal is a Senior Principal Research Scientist at Honeywell Laboratories, Minneapolis. He received his Ph.D. from Carnegie Mellon University, and B.Tech. from Indian Institute of Technology, Bombay. His areas of expertise include large-scale and dynamic optimization, hybrid systems, process modeling and simulation, theory of differential algebraic equations, non-smooth systems, and multivariable control. He has published extensively and chaired and organized conference sessions and symposia on these topics. Currently, Dr. Gopal is co-PI on

Intrusion Tolerance via Multimodel Predictive Control (DARPA/ITS) program and a key investigator on the Real-Time Adaptive Resource Management (DARPA/QUORUM) program. He has also led algorithm development efforts in DARPA-sponsored AICE, SEC and JFACC programs. At Honeywell, Dr. Gopal has been PI on the following IR&D projects—Optimization of Hybrid Systems; Uncertainty in Process Computations; and Multi-Aircraft Conflict Resolution in Free Flight. He is also a contributor to the *Dept. of Energy Technology Roadmap for the US Petroleum Industry (2000, Draft)* and the *2002 McGraw Hill Yearbook of Science and Technology.*

David Gross

David C. Gross is a software systems engineer with Modeling and Simulation Technology within the Boeing PhantomWorks. He has conducted applied research in the areas of software development process, software reuse/re-engineering, and software quality, as they relate to simulation.

He is currently involved in applied research in advanced simulation technologies such as knowledge-based simulation, visualization, and graphical user interfaces. Mr. Gross holds a Bachelor of Science in Computer Science/Engineering from Auburn University and a Master of Operations Research at the University of Alabama at Huntsville. Mr. Gross is a doctoral candidate at the University of Central Florida.

Scott Harmon

Scott Harmon is currently CEO and Chief Scientist for Zetetix. During his career, he has gained a wide range of experience in government, academia and industry where he served in both technical and management roles. He has contributed to the areas of robotics, intelligent systems, modeling and simulation, systems engineering, information system architectures, human behavior representations, and simulation validation. He is presently developing tools for modeling complex information systems and a physical theory that explains the behavior of all information systems.

Mr. Harmon has a BS in Physics and an MS in Mechanical Engineering, both from Arizona State University. He has written over 70 technical publications on robotics, systems engineering, human behavior representations, and modeling and simulation.

Samuel Nicholson

Samuel Nicholson graduated from the U. S. Naval Academy in 1966 with a BS in Marine Engineering. After attending nuclear power and submarine training, he served in the submarine force for the balance of his 21 year Navy career. This culminated in a senior staff position with the Commander Submarine Forces Pacific following his 4 year command tour of the USS James K. Polk (SSBN 645) (Blue). Upon retiring from the Navy, he worked in the professional services field as a consultant to Navy and commercial organizations. Subsequently he has

provided technical and programmatic staff support to various submarine, automated design, and information system programs at the Defense Advanced Research Projects Agency (DARPA).

Hassen Saidi

Dr. Hassen Saidi holds a masters in computer science from the University of Paris 7 and a Ph.D. in computer science from the University of Grenoble, France. His main research interest is program analysis and verification using mathematical tools. He is interested in applying algorithmic, deductive, and compositional methods to analyze and verify complex software systems. He worked on the combination of theorem proving and model checking through abstract interpretation theory to reduce the complexity of large systems. He is a computer scientist in the System Design Laboratory at SRI since May 1999.

Tommaso Toffoli

Prof. Tommaso Toffoli received a Doctorate in Physics from the University of Rome, Italy, in 1967, and a Ph.D. in Computer and Communication Science from the University of Michigan, Ann Arbor, in 1976. In 1977 he joined the MIT Laboratory for Computer Science as Research Scientist, where he became Principal Research Scientist and leader of the Information Mechanics Group in 1986. In 1995 he joined the faculty of the Boston University Electrical and Computer Engineering Department, as Associate Professor. He is senior member of the IEEE, and editorial board member of Complex Systems and the InterJournal.

He has been studying physical aspects of information at the microscopic and macroscopic levels. He is active in reversible computing, fine-grained parallel processing, and connections between quantum structure and information theory.

WORKSHOP PRESENTATIONS

All Roads Lead to Rome

Prof. Tommaso Toffoli
Boston University

Abstract

From the industrial revolution engineers inherited an overwhelming concern for optimal behavior, efficient use of resources, productiveness---in sum, to seek the best. In this quest, physics has been the main source of inspiration and examples as well as an indispensable provider of concepts and techniques.

But, in real life, it is not those who seek THE BEST that get the prize; to stay in the race you need only look for the GOOD ENOUGH---but make DAMN SURE you get it. What are the mathematics and physics of this less naïve ``evolutionary strategy?"

We'll begin with a catalog of conventional ``optimal design myths;" these will be used as strawmen as we proceed to outline a sounder ``natural theology." We'll then argue that information assurance is basically the art of constructing systems where by design the overwhelming majority of trajectories lead to an acceptable outcome---where ``all roads lead to Rome". In this sense, information assurance is best viewed as a statistical mechanics of higher order.

References

Dennis Geller, "A Socratic Dialogue," Datamation, 20(11), November 1974, pp72-75.

Edwin Jaynes, "Information Theory and Statistical Mechanics I," Physical Review, 106(4), 1957, pp620-630.

Edwin Jaynes, "Information Theory and Statistical Mechanics II," Physical Review, 108(2), 1957, pp171-190.

R.W. Keyes & Rolf Landauer, "Minimal Energy Dissipation in Logic," IBM Journal of Research and Development, 14, 1970, pp152-157.

Walter Kirchherr, Ming Lee & Paul Vitanyi, "The Miraculous Universal Distribution," The Mathematical Intelligencer, 19(4), 1997, pp7-15.

Tommaso Toffoli, "Action, or the Fungibility of Computation," Feynman and Computation: Exploring the Limits of Computation, A.J.G. Hey, ed., Perseus Books, Reading, MA, 1999, pp349-392.

Tommaso Toffoli, "What You Always Wanted to Know about Genetic Algorithms but Were Afraid to Hear," Festschrifts in Honor of John Holland, Lashon Booker, Stephanie Forrest, Melanie Mitchell & Rick Riolo, eds., Center for Studies in Complex Systems, University of Michigan, Ann Arbor, MI, June 1999, pp131-136.

Information Assurance Mathematics

Dr. Lee Benzinger
NAI Laboratories

Abstract

Critical systems have traditionally achieved limited Information Assurance (IA) through special purpose development and through isolation from the rest of the world. The need for IA is rapidly increasing, but, at the same time, new factors have greatly increased the difficulty of achieving meaningful IA. These factors include the fact that today's systems are dramatically more complex and interconnected with other systems that span a variety of IA capabilities and that it is typical for cost reasons to employ COTS components in such systems. COTS components provide functionality but usually have very weak IA properties. To transform building systems with critical IA properties from an art to an engineering discipline, a scientific basis for system prediction, analysis, and reproducibility is needed. Such a scientific basis requires a mathematics for reasoning about systems and their IA properties in (preferably) closed form expressions.

Under DARPA contract, NAI Labs is developing a body of mathematics (Information Assurance Mathematics -- IAM) for analyzing systems, components, IA policies, and their compositions. This work builds upon and extends prior research in the mathematical fundamentals of ultra-large-scale systems known as the (Loral Western Development Labs) WDL theory. These fundamentals include a large number of proved theorems governing ultra-large-scale system component interactions. IAM is intended to model components, component policies, and their composition to form systems and system polices. The IAM approach includes capturing the subtle interactions that are possible when components and policies are combined. Our research is focusing on two technical objectives for developing IAM:

- How to complete the existing mathematical theory as a mathematical basis for systems modelling; and
- How to extend this mathematical systems theory to a mathematical theory for modelling software systems.

To complete the existing theory, our research will investigate operators on components and policies and properties of combinations of such components and policies under these operators. In particular, there are two properties, closure and transitivity, that are important for describing

IA properties of combinations of components and policies. To extend the mathematical systems theory to a mathematical theory for modeling software systems we will investigate the properties of mappings from systems level concepts to software level concepts. We expect that the operators on systems models will need to be extended to operators on software models. We also expect that the systems properties of closure and transitivity will need to be appropriately extended to concepts that are meaningful for software properties of components and their policies.

Our intent is that IAM provide the theoretical basis for a systems and software systems design and analysis tool. The target tool users would be engineers, systems and software systems architects.

References

Martin Abadi & Leslie Lamport, Composing Specifications, Technical Report 66, Digital Systems Research Center, Palo Alto, California, October 1990.

L. Benzinger, "Applying the WDL Composition Theory to Analyze Security Architectures," Proc. Composition Software Architecture Workshop, Monterey, CA, January 1998, np.

G.W. Dinolt, L. A. Benzinger & M.G. Yatabe, "Combining Components and Policies," Proc. Computer Security Foundations Workshop VII, IEEE Computer Society Press, 1994, np.

Joshua Guttman, "Filtering Postures: Local Enforcement of Global Policies," Proc. 1997 IEEE Symposium on Security and Privacy, IEEE Computer Security Press, pp120-129.

Dale M. Johnson and E. Javier Thayer, "Security and the Composition of Machines," Proc. 1988 Computer Security Foundations Workshop, IEEE Computer Society Press, June 1988, pp72-89

H.M. Hinton, "Composing Partially-Specified Systems," Proc. 1998 IEEE Symposium on Security and Privacy, IEEE Computer Society Press, May 1998, pp27-37.

Daryl McCullough, "Noninterference and the Composability of Security Properties," Proc. 1988 IEEE Symposium on Security and Privacy, IEEE Computer Society Press, May 1988, pp177-186.

John McLean, "A General Theory of Composition for Trace Sets Closed under Selective Interleaving Functions," Proc. 1994 IEEE Symposium on Security and Privacy, IEEE Computer Society Press, 1994, np.

John Rushby, "Composing Trustworthy Systems: A Position Paper," Proc. NATO RSG2 Working Conference on Composability, October 1991, np.

Nature of Interactions and Emergent Properties

Dr. Andrew Belyavin
DERA Centre for Human Sciences

Abstract

Understanding the macroscopic behaviour of information systems is seen as equivalent to the analysis of any form of human collective behaviour, and thus depends on the nature of the behaviour of the individual entities and the interactions between them.

It is a familiar phenomenon in the world of physics that the nature of the interactions between the entities in a system can have a substantial impact on the macroscopic properties of the system. This is most easily demonstrated for systems composed of similar or identical entities with standardised patterns of interaction. Similar phenomena have been found for simple systems of interactions between human players such as traffic flow systems, where the nature of the interaction produces clear observable macroscopic phenomena. As the nature of the interactions become more complex and the structure of the interactions becomes more defined, the direct relationship between emergent properties and the nature of the interactions appears more blurred. For highly organised - even hierarchical - command and control systems, the properties of individual entities can dominate the nature of the interactions, and the macroscopic properties of the system no longer emerge from the nature of the interactions between the entities.

The talk will not propose a specific model of how interactions determine macroscopic properties, but will draw attention to the way different structures appear to relate to macroscopic properties from observation, and therefore the need to consider a range of approaches to isolating the phenomena under discussion.

Categorical Modeling of Information Assurance

Prof. Thomas L. Clarke & Dr. Dennis K. McBride
University of Central Florida

Abstract

First developed to address problems in topology, category theory has developed into a general mathematical theory of structures and systems of structures. Category theory provides a unifying and compact mathematical modeling language that enables the mapping via functors of problems from one area of mathematics to another where the solution may be easier to find. Computer science has also adopted category theory to applications in the field of algebraic semantics, the theory of programming languages, and automata theory.

There is a correspondence between the functors of category theory and the exchange of information between systems. Philosophically, category theory can be seen as providing a theory of concepts in that the categorical structure of information constitutes a concept. A functor that faithfully maps information between the categorical structures of two systems is thus mapping the concepts of the two systems. Checking for dissonance between the mapped concept and the prior categorical structure can provide a test for information assurance. Category theorems proving when functors are full and faithful have been applied to the field of training simulator interoperability, but can be extended easily to other informational contexts.

A particularly fruitful area for the application of category theory to information is in providing functors or maps between the wide variety of logics available; in addition to classical logic, there are fuzzy logics, affine logics, quantum logics, linear logic etc. Of these, linear logic may be more universal than classical logic. Linear logic shows much promise for the modeling of deep structures in quantum mechanics and as a possible logic for the brain. Pratt in particular proposes constructing the world out of the fabric of Chu spaces which are k-valued binary relations from a set to the Boolean algebra on the set Pratt was able to prove that Chu spaces are capable of expressing all of mathematics in a categorical sense. This connection with logic assures that category theory can be applied using fairly standard inference engine techniques running on ordinary computer hardware.

The talk will introduce category theory and will discuss how it applied to problems of information systems assurance as well as training simulator validation.

References

Jiri Adamek, Horst Herrlich & George E. Strecker , Abstract and Concrete Categories: The Joy of Cats, John Wiley & Sons, New York, NY, 1990.

Michael Barr & Charles Wells, Category Theory for Computer Science, Prentice Hall , New York, NY, 1995.

Bart Jacobs, Categorical Logic and Type Theory, Elsevier Science Publishers, Amsterdam, The Netherlands, 1999.

Horst Herrlich & Hans-E. Porst, eds., Category Theory at Work, Heldermann Verlag, Berlin, Germany, 1991.

Saunders MacLane, Categories for the Working Mathematician, Springer, New York, NY, 1998.

Benjamin C. Pierce, Basic Category Theory for Computer Scientists, MIT Press, Cambridge, MA, 1991.

David Pitt et al., eds., Proc. Category Theory and Computer Programming Tutorial and Workshop, Guildford, U.K., September 16-20, 1985, Springer-Verlag, Berlin, Germany, 1986.

Cyberlogic: A Logical Characterization of Macroscopic Phenomena of Complex Information Systems

Dr. Hassen Saidi
SRI International

Abstract

The world is rapidly moving to an information-based economy where large volumes of information are stored, communicated, and traded in electronic form. This information infrastructure includes medical records, financial information, electronic cash, forms of identification, credit card numbers, phone conversations, video, and virtually everything else of value. In such an information-centric world, even small vulnerabilities can lead to severe damage and loss. These days the news media are filled with accounts of electronic break-ins or security flaws in popular software systems, and the prospects are very bleak for the future of information security. A critical question for such an information infrastructure is: "How can we ensure the fidelity and privacy of all this information?" On the one hand, this requires carefully engineered protocols and information access-control mechanisms, and intrusion detection and tracking mechanisms. On the other hand, we need to understand the mathematical foundations that can be used to model and analyze the security properties of information-based transactions, and the treats they can be subject to.

The role of cyberlogic is to provide abstract yet accurate models of threats and countermeasures. In cyberlogic we will model systems, their properties and how they are subject to human and physical attacks. Cyberlogic will also model countermeasures that can be based on the physical or behavioral characteristic of the context and the attacker. We will also propose models for human understanding of security in order to detect mismatches between the actual security level of a system and human conception of what the security of system is suppose to provide. We will elaborate a collection of scenarios where security is potentially compromised by several means. We will model in cyberlogic the scenarios as well as the adequate countermeasures.

References

M. Abadi, M. Burrows, B. Lampson & G. Plotkin, "A Calculus for Access Control in Distributed Systems," ACM Transactions on Programming Languages and Systems, 15, 1993, np.

Garett Birkoff, "The Role of Modern Algebra in Computing," Computers in Algebra in Number Theory, American Mathematical Society, 1971, np.

Iliano Cervesato, Nancy A. Durgin, Patrick Lincoln, John C. Mitchell & Andre Scedrov, "A Meta-Notation for Protocol Analysis," Proc. 12th IEEE Computer Security Foundations Workshop (CSFW'99), Mordano, Italy, June 1999, np.

Paul Cosher, Joshua Jaffe & Benjamin Jun, "Differential Power Analysis," Proc. 19th Annual International Cryptography Conference "Advances in Cryptography" (Crypto'99), LNCS 1666, Santa Barbara, CA, 1999, np.

Manuel Clavel, Francisco Duran, Steven Eker, Jose Meseguer & Mark-Oliver Stehr, "Maude as a Formal Meta-Tool," Proc. FM'99, The World Congress on Formal Methods in the Development of Computing Systems, Toulouse, France, September 20-24, 1999, np.

K. M. Chandy & J. Misra, "How Processes Learn," Journal of Distributed Computing, 1, 1986, pp40-52.

S. Dawson, S. DeCapitani, P. Lincoln & P. Samarati, "Minimal Data Upgrading to Prevent Inference and Association Attacks," Proc. 18th ACM Principles of Database Systems, Philadelphia, PA, 1999, np.

D. Dolev & A. Yao, "On the Security of Public-Key Protocols," IEEE Transactions on Information Theory, 2(29), 1983, np.

Ronald Fagin, Joseph Y. Halpern, Yoram Moses & Moshe Y. Vardi, Reasoning About Knowledge, MIT Press Books, Cambridge, MA, 1995.

Jim Gray, What Next? A Dozen Remaining IT Problems, My Turing Award Lecture," ACM Turing Award at the ACM Awards Banquet, New York, NY, April 15, 1999.

J. Halpern & Y. Moses, "Knowledge and Common Knowledge in a Distributed Environment," Journal of the ACM, 3(37), 1990, np.

Patrick Lincoln, John Mitchell, Mark Mitchell & Andre Scedrov, "A Probabilistic Polytime Framework," 5th ACM Conference on Computer and Communications Security (CCS'98), San Francisco, California, November 5, 1998, np.

P.D. Lincoln, J.C. Mitchell, M. Mitchell & A. Scedrov. "Probabilistic Polynomial-time Equivalence and Security Protocols," FM'99 World Congress On Formal Methods in the Development of Computing Systems, Toulouse, France, September, 1999, np.

Gavin Lowe, "Breaking and Fixing the Needham-Schroeder Public-Key Protocol using FDR," Tools and Algorithms for the Construction and Analysis of Systems, Margaria and Steffen (eds.), volume 1055 of Lecture Notes in Computer Science, Springer Verlag, 1996, pp147-166. Also in Software Concepts and Tools, 17, 1996, pp93-102.

M.S. Mahoney, "The Structures of Computation," Proc. International Conference on the History of Computing, Paderborn, Germany, 1998, np.

M.S. Mahoney, "The History of Computing in the History of Technology," Annals of the History of Computing, 10, 1988, pp113-125.

John K. Millen, "A Necessarily Parallel Attack," Proc. Workshop on Formal Methods and Security Protocols (FMSP'99), Trento, Italy, July 1999, np.

David Monniaux, "Decision Procedures for the Analysis of Cryptographic Protocols by Logics of Belief," Proc. 12th IEEE Computer Security Foundations Workshop (CSFW'99), Mordano, Italy, June 1999, np.

P.G. Neumann, Computer-Related Risks, Addison-Wesley/ACM Press, 1995.

P.G. Neumann, Practical Architectures for Survivable Systems and Network Final Report, Phase One, Project 1688, SRI International, Menlo Park, CA, January 1999.

John Rushby, "Using Model Checking to Help Discover Mode Confusions and Other Automation Surprises," Proc. 3rd Workshop on Human Error, Safety, and System Development (HESSD'99), Liege, Belgium, 7-8 June 1999, np.

N. Shankar, S. Owre, J.M. Rushby & D.W.J. Stringer-Calvert, PVS Prover Guide, Computer Science Laboratory, SRI International. September 1999.

Is Fisher Information a Governing Concept in Cyberspace?

Dr. Roger Cox
Sandia National Laboratories

Abstract

In a series of physics journal articles in the late 1990's, culminating in a Cambridge University Press textbook published in 1999, Roy Frieden of the University of Arizona developed a unifying methodology deriving most of the governing equations of physics from a single statistical principle - that of Fisher Information. These include Maxwell equations, Klein-Gordon equation, Boltzmann law, Schroedinger wave equation (time-invariant), $E=mc^2$, Heisenberg uncertainty principle, and certain quantum gravity equations. Much of theoretical physics can thus be thought as derivable from the mathematical statistics of continuous space.

Frieden's derivations spring from the assumption that a physical law is the optimal expression of the information available from any experimentation. Fisher information becomes the key concept, since it is the minimum uncertainty that must surround any prediction that is a function of observed data. Alternatively, it provides a lower bound for the variance associated with any statistical estimator, as expressed mathematically by the Cramer-Rao inequality. Fisher information is expressed as the variance of the natural log of the derivative of a conditional density function. Frieden then proposes two Extreme Physical Information (EPI) principles that describe how Fisher information is manipulated to provide governing equations.

This investigation explored the feasibility of applying Frieden's mathematical machinery to problems on discrete and hybrid discrete/continuous spaces. Frieden's work is limited to continuous spaces, since that is where most classical physics operates. Cyberspace is better described by discrete sets or discrete/continuous hybrids such as networks. To address problems

in discrete/hybrid spaces using Frieden's machinery, we need to extend it to discrete sets and hybrid spaces, e.g. where the Omega space of the probability space is the set of possible network designs.

This involves extending Fisher Information, an inherently local measure, to sets that only allow partial preorderings, rather than strict orderings. This is nontrivial since arbitrary discrete sets don't allow a local information measure such as Fisher's, and must instead use global entropy measures such as Shannon's. We develop a path metric on the lattice induced by the preordering to extend the Cramer-Rao result to discrete sets. The result is a "generalization of generalizations" of Cramer-Rao. Specific cases of our discrete Fisher information are famous generalizations of Cramer-Rao: Fraser-Guttman, Keifer, Chapman-Robbins, Bhattacharyya.

The resulting discrete Fisher information measure can then be used directly in Frieden's mathematics. We show how such a measure can be used in Frieden's method to derive the discrete set equivalent of the famous equation E=mc2.

The final step, prior to applying Frieden's work with our Fisher information measure to problems in cyberspace, is to identify where it is appropriate to assume Frieden's EPI principles are valid. We propose an approach to do this by formally developing classical statistics using EPI and, through this exercise, identifying what problems EPI can address that are outside of the scope of classical statistics and what modelling limits are valid for both statistics and EPI.

References

B.R. Frieden, Physics from Fisher Information: A Unification, Cambridge University Press, Cambridge, United Kingdom, 1998.

B.R. Frieden, "Estimation of Distribution Laws, and Physical Laws, by a Principle of Extremized Physical Information," Phys. A, 198, 1993, pp272-338.

M. Reginatto, "Derivation of the Equations of Nonrelativistic Quantum Mechanics Using the Principle of Minimum Fisher Information," Phys. Rev. A, 58(3), 1998, pp1775-1778.

A. Puente, M. Casa & A. Plastino, "Fisher Information and Semiclassical Methods," Phys. Rev. A, 59(5), 1999, pp3211-3217.

W.J. Cocke, "Incompressible Turbulence and Minimum Fisher Information: Correlations between Velocity and Pressure," Phys. Fluids, 10(8), 1998, pp2055-2060.

H. Cramer, Mathematical Methods of Statistics, Princeton University Press, Princeton, 1961.

C.R. Rao, Linear Statistical Inference and Its Applications, Wiley, New York, 1973.

P.J. Bickel & K.A. Doksum, Mathematical Statistics: Basic Ideas and Selected Topics, Holden-Day, San Francisco, 1977.

D.G. Chapman & H. Robbins, "Minimum Variance Estimation without Regularity Assumptions," Ann. Math. Statist., 22, 1951, pp581-586.

A. Bhattacharyya, "On Some Analogues to the Amount of Information and their Uses in Statistical Estimation," Sankhya, 8, 1946, pp1-14.

D.A.S. Fraser & I. Guttman, "Bhattacharyya Bounds without Regularity Assumptions," Ann. Math. Statist., 23, 1952, pp629-632.

J. Keifer, "On Minimum Variance Estimation," Ann. Math. Statist., 23, 1952, pp627-629.

S. Kullback, Information Theory and Statistics, Wiley, New York, 1959.

H. Risken, The Fokker-Planck Equation, Springer-Verlag, Berlin, 1984.

A.J. Stam, "Some Inequalities Satisfied by the Quantities of Information of Fisher and Shannon," Information and Control, 2, 1959, pp101-112.

Kolmogorov Complexity as a Conserved Parameter of Information Assurance

Mr. Scott Evans & Dr. Steven Bush
GE Corporate Research and Development

Abstract

Kolmogorov Complexity is a fundamental attribute of information that introduces relationships that further develop the laws of information assurance. This paper reviews the concepts of Kolmogorov Complexity and promotes this parameter for use in conservation within physics of information. Analogies between thermodynamics, Kolmogorov Complexity, and the problem of information assurance are explored. A law of conservation of data and complexity is proposed. Finally, opportunities are presented to exploit these laws to achieve a better understanding of information assurance.

References

G.J. Chaitin, "The Limits of Mathematics --- Course Outline and Software", Lecture Notes in Computer Science, 888, 1995, p1.

G.J. Chaitin, A. Arslanov & C. Calude, "Program-size Complexity Computes the Halting Problem", Bulletin of the European Association for Theoretical Computer Science, 57, October 1995, pp198—200.

G.J. Chaitin, "Randomness in arithmetic and the decline and fall or reductionism in pure mathematics", Bulletin of the European Association for Theoretical Computer Science, 50, June 1993, pp314—328.

Ming Li & Paul Vitanyi, Introduction to Kolmogorov Complexity and its Applications, Springer-Verlag, New York, NY, 1993.

Manfred Denker, W. A. Woyczynski & Bernard Ycart, Introductory statistics and random phenomena: uncertainty, complexity, and chaotic behavior in engineering and science, Birkhauser Boston Inc., Cambridge, MA, USA, 1998.

Douglas C. Giancoli,. General Physics, Prentice Hall, Inc., Englewood Cliffs, NJ, nd.

P. Fraundorf, "Heat Capacity in Bits," April 28, 2000, http://www.umsl.edu/~fraundor/ifzx/cvinbits.html. An active revision of cond-mat/9711074 in the Los Alamos archives.

T.M. Cover & J. A. Thomas, Elements of Information Theory, John Wiley & Sons, New York, NY, 1991.

W.H. Zurek, ed., Complexity, Entropy, and the Physics of Information, Santa Fe Institute, Santa Fe, NM, USA, 1989.

Rolf Herken, The Universal Turing Machine, A Half-Century Survey, Springer-Verlag, New York, NY, 1995.

A Physical Model of the Behavior of Information Systems

Mr. Scott Harmon
Zetetix

Abstract

The model of information systems, discussed in this presentation, builds upon past work in the physics of computation, computational complexity theory, and information theory. This model represents the impact that existing physical laws have upon the state and behavior of objects in the abstract worlds that information systems create, maintain, store and communicate. It posits that information exists only as modulated energy quantized into abstract symbols. Two forms of these symbols exist, pure data and instructions. The devices in information systems can execute instruction symbols. This execution process transforms pure data inputs into pure data outputs. All information flows through processing, communication links and memory require the devices of an information system to perform physical work. This property suggests that the work required to support an information flow is proportional to the rate of that flow times the resistance of the device supporting it. This model also suggests that the existence of a particular type of state information, as goals, in a system can drive the information flows within that system. The rate of goal-driven information flow is proportional to the potential of the driving goal divided by the device resistance.

Another form of information flow is analogous to diffusion. The rate of information diffusion within a system is proportional to a diffusion constant times the gradient of information complexity. Information complexity is proportional to the number of executable dependencies

that could exist between the different abstract property states represented within that system. The observability of both information flow rates and content complexity permits the calculation of diffusion constants for different information systems. The similarities between highly complex systems and disordered systems suggests that system complexity, and therefore information complexity, is proportional to the physical entropy of a system. Since information contributes to system complexity, it also lends components to system entropy. This linkage between the entropy of a system and the information that could influence that system's behavior permits extending the Second Law of Thermodynamics to explain such phenomena as information leakage and progressive resource saturation. These two phenomena are commonly observed in complex information systems of today.

References

R.B. Ash, Information Theory, Dover Publications, New York, NY, 1965.

S. Barta, "Relation between Information and Thermodynamic Entropy," J. of Electrical Engineering, **48** (7-8), 1997, pp169-174.

J.D. Bekenstein, "Entropy Content and Information Flow in Systems with Limited Energy," Physical Review D, **30**, 1984, pp1669-1679.

C.H. Bennett, "Logical Reversibility of Computation," IBM J. of Research & Development, **17**, 1973, pp525-532.

C.H. Bennett, "The Thermodynamics of Computation - A Review," Int. J. of Theoretical Physics, **21** (12), 1982, pp905-940.

C.H. Bennett, "Logical Depth and Other Algorithmically Defined Properties of Finite Objects," Proc. 2nd IEEE Workshop on Physics & Computation, Dallas, TX, 1993, IEEE Computer Society Press, Los Alamitos, CA, pp75-77.

C.H. Bennett & R.W. Landauer, "The Fundamental Physical Limits of Computation," Scientific American, **253** (1), 1985, pp48-56.

L. Brillouin, "Maxwell's Demon Cannot Operate: Information and Entropy," Int. J. of Applied Physics, **22**, 1950, pp334-337.

L. Brillouin, Science and Information Theory, Academic Press, New York, NY, 1956.

G.J. Chaitin, "Algorithmic Information Theory," IBM J. of Research & Development, **21**, 1977, pp 350-359.

J.P. Crutchfield & C.R. Shalizi, "Thermodynamic Depth of Causal States: Objective Complexity via Minimal Representations," Physical Review E, **59**, 1999, pp 275-283.

M. Davis, Computability and Unsolvability, Dover Publications, Inc., New York, NY, 1982.

P.A. Dufort & C.J. Lumsden, "The Complexity and Entropy of Turing Machines," Proc. 3rd IEEE Workshop on Physics & Computation, Dallas, TX, 1994, IEEE Computer Society Press, Los Alamitos, CA, pp227-232.

R.P. Feynman, Feynman Lectures on Computation, Addison-Wesley Publishing Co., Inc., Reading, MA, 1996.

E. Fredkin & T. Toffoli, "Conservative Logic," Int. J. of Theoretical Physics, **21** (3/4), 1982, pp219-253.

N. Fuchikami, H. Iwata & S. Ishioka, "Thermodynamic Entropy of Computer Devices," J. Physical Soc. of Japan, **68** (12), 1999, pp3751-3754.

D. Gabor, "Communication Theory and Physics," Philosophy Mag., **41**, 1950, pp 1161-1187.

P. Gacs, "The Boltzmann Entropy and Randomness Tests," Proc. 3rd IEEE Workshop on Physics & Computation, Dallas, TX, 1994, IEEE Computer Society Press, Los Alamitos, CA, pp 209-216.

N.A. Gershenfeld, "Signal Entropy and the Thermodynamics of Computation," IBM Systems J., **35** (3/4), 1996, pp 577-586.

D.C. Gross et al., "Report from the Fidelity Implementation Study Group," Paper No. 99S-SIW-167, Proc. of 1999 Spring Simulation Interoperability Workshop, Orlando, FL, March 1999, np.

E. Gurari, An Introduction to the Theory of Computation, Computer Science Press, Rockville, MD, 1989.

S.Y. Harmon, "Evaluating and Comparing Information Systems," Proc. of the 1998 IEEE International Conf. on Systems, Man, and Cybernetics, San Diego, CA, October 1998, np.

S.Y. Harmon, "Evaluating the Performance of Intelligent Systems," Proc. of the 3rd World Multiconference on Systemics, Cybernetics and Informatics, Orlando, FL, July 1999, np.

A.J.G. Hey, ed., Feynman and Computation, Perseus Books Publishing, Reading, MA, 1999.

E.T. Jaynes, "Information Theory and Statistical Mechanics I," Physical Review A, **106** (4), 1957, pp 620-630.

E.T. Jaynes, "Information Theory and Statistical Mechanics II," Physical Review A, **108** (2), 1957, pp 171-190.

F.W. Kantor, Information Mechanics, John Wiley & Sons, New York, NY, 1977.

R.W. Keyes, "Fundamental Limit in Digital Information Processing," Proc. IEEE, **69**, 1981, pp 267-278.

A.N. Kolmogorov, "Three Approaches to the Quantitative Definition of Information," Problems of Information Transmission, **1** (1), 1965, pp 1-7.

R.W. Landauer, "Irreversibility and Heat Generation in the Computing Process," IBM J. of Research & Development, **5**, 1961, pp183-191.

R.W. Landauer, "Information Is Physical," Physics Today, **44** (5), 1991, pp23-29.

D.S. Lebedev & L.B. Levitin, "The Maximum Amount of Information Transmissible by Electromagnetic Field," Soviet Physics Doklady, **8**, 1963, pp 377-379.

D.S. Lebedev & L.B. Levitin, "Information Transmission by Electromagnetic Field," Information & Control, **9** (1), 1966, pp 1-22.

H.S. Leff & A.F. Rex, Maxwell's Demon: Entropy, Information, Computing, Princeton University Press, Princeton, NJ, 1990.

L.A. Levin, "Various Measures of Complexity for Finite Objects (Axiomatic Description)," Soviet Mathematics Doklady, **17** (2), 1976, pp 522-526.

L.B. Levitin, "Physical Information Theory Part I. Quasiclassical Systems," Proc. 2nd IEEE Workshop on Physics & Computation, Dallas, TX, 1993, IEEE Computer Society Press, Los Alamitos, CA, pp 210-214.

L.B. Levitin, "Physical Information Theory Part II. Quantum Systems," Proc. 2nd IEEE Workshop on Physics & Computation, Dallas, TX, 1993, IEEE Computer Society Press, Los Alamitos, CA, pp 215-219.

M. Li & P.M.B. Vitanyi, "Theory of Thermodynamics of Computation," Proc. 2nd IEEE Workshop on Physics & Computation, Dallas, TX, 1993, IEEE Computer Society Press, Los Alamitos, CA, pp42-46.

M. Li & P.M.B. Vitanyi, An Introduction to Kolmogorov Complexity and Its Applications, 2nd ed., Springer-Verlag, New York, NY, 1997.

S. Lloyd, "Physical Measures of Complexity," 1989 Lectures in Complex Systems, E. Jen, ed., Addison-Wesley, Redwood City, CA, 1990, pp 67-73.

J. Machta, "Entropy, Information, and Computation," American J. of Physics, **67** (12), 1999, pp1074-1077.

N.H. Margolus, "Fundamental Physical Constraints on the Computational Process," Nanotechnology: Research and Perspectives, B.C.Crandall & J. Lewis, eds., MIT Press, Cambridge, MA, 1992, np.

P. Martin-Lof, "The Definition of Random Sequences," Information & Control, **9** (6), 1966, pp 602-619.

D. Matzke, ed., Proc. of the 1992 Workshop on Physics and Computation (PhysComp '92), 2-4 October 1992, Dallas, TX, USA, IEEE Computer Society Press, Los Alamitos, CA, 1992.

D. Matzke, ed., Proc. of the 1994 Workshop on Physics and Computation (PhysComp '94), 17-20 November 1994, Dallas, TX, USA, IEEE Computer Society Press, Los Alamitos, CA, 1994.

C.H. Papadimitriou & M. Sipser, "Communication Complexity," J. of Computer & System Sciences, **28** (2), 1984, pp 260-269.

P.L. Patterson, "Entropy, Fault Tolerance, and Multicomputer Networks," Proc. 2nd IEEE Workshop on Physics & Computation, Dallas, TX, 1993, IEEE Computer Society Press, Los Alamitos, CA, pp232-236.

N. Pippenger, "Complexity Theory," Scientific American, **239**, 1978, pp 90-100.

R.P. Poplavskii, "Thermodynamic Models of Information Processes," Soviet Physics Uspekhi, **115**, 1975, pp 222-241.

C.E. Shannon & W. Weaver, The Mathematical Theory of Communication, Univ. of Illinois Press, Urbana, IL, 1963.

R.J. Solomonoff, "The Discovery of Algorithmic Probability," J. of Computer & System Sciences, **55** (1), 1997, pp 73-88.

T. Stonier, "Towards a General Theory of Information: Information and Entropy," Future Computing Systems, **2** (4), 1990, pp409-427.

L. Szilard, "On the Decrease of Entropy in a Thermodynamic System by the Intervention of Intelligent Beings," Quantum Theory and Measurement, J.A. Wheeler & W.H. Zurek, eds., Princeton University Press, Princeton, NJ, 1983, pp 539-548.

T. Toffoli, M. Biafore & J. Leao, eds., Proc. of the 1996 Workshop on Physics and Computation (PhysComp '96), 22-24 November 1996, Boston, MA, USA, New England Complex Systems Institute, Cambridge, MA, 1996.

P.M.B. Vitanyi, "Multiprocessor Architectures and Physical Law," Proc. 3rd IEEE Workshop on Physics & Computation, Dallas, TX, 1994, IEEE Computer Society Press, Los Alamitos, CA, pp 24-29.

W. Weaver, "Science and Complexity," American Scientist, **36**, 1968, pp 536-544.

D.H. Wolpert, "Memory Systems, Computation, and the Second Law of Thermodynamics," Int. J. of Theoretical Physics, **31**, 1992, pp 743-785.

W.H. Zurek, "Algorithmic Randomness and Physical Entropy," Physical Review A, **40** (8), 1989, pp4731-4751.

A.K. Zvonkin & L.A. Levin, "The Complexity of Finite Objects and the Development of the Concepts of Information and Randomness by Means of the Theory of Algorithms," Russian Mathematical Surveys, **25** (6), 1970, pp 83-124.

Synopsis of Current Directions in Cartography of Cyberspace

Mr. Michael Frentz
BBN Technologies

Abstract

One of the goals of this task is to develop a visually-based extensible research framework – a visual taxonomy – to enable collaborative investigation into the security aspects of networks and the mapping of security concepts to network implementation. A taxonomy is not a neutral structure – its organization implicitly creates a theory for the problem space and will determine what data gets collected and what questions can reasonably get answered.

Security is a bit of an odd discipline – it is more akin to the insurance industry or the automotive safety industry – than a typical "positive" engineering discipline. Security only provides explicit value in the event that something goes wrong which will prevent a mission from being accomplished (with an implicit appropriate performance measure). Security is always peripheral to the primary mission for which a system exists and always adds cost in some dimension. A user base – especially one under performance pressure – will prefer to do without it if they do not perceive a favorable cost/benefit tradeoff.

We have reviewed about a dozen primary papers on security-relevant taxonomies. While each presents a valid specific decomposition for a particular sub-area, there also tends to be somewhat of a disjointedness in the classes of the presented concepts. We attribute the inadequacy due to the mixing of concepts from logically disparate domains and propose to view a system in terms of three distinct but interrelated worldviews – mission oriented, functionally oriented, and implementation based.

The worldviews provide an interesting insight into security flaws – while attacks always occur in the third worldview, they are only felt and are measurable in the first worldview. If the user base can still accomplish its mission successfully, the attack is by definition benign. The lack of traceability across worldviews is a primary contributor of system insecurities.

Re-expressing system concepts in a layered worldview may allow "locally complete" descriptions of systems to be generated for the multiple quasi-independent domains. Standardization of a nomenclature for layer 2 (functional view) would appear to be a necessary prerequisite for the development of secure system standards for COTS software. A long-term goal to financially encourage COTS vendors to produce that level of documentation in order to effectively sell to the government or "critical private" systems market should be feasible and would be beneficial to increasing the security accountability of widely used systems.

References

J.D. Howard & T.A. Longstaff, A Common Language for Computer Security Incidents, Sandia Report No. SAND 98-8667, Sandia National Laboratories, Albuquerque, NM, USA, October 1998.

J.D. Howard, "An Analysis of Security Incidents on the Internet, 1989-95," Ph.D. Dissertation, Carnegie-Mellon University, Pittsburg, PA, 1996, www.cert.org/research/JHThesis/start.html.

C. Landwehr, A. Bull, J. McDermott & W. Choi, "Taxonomy of General Computer Program Security Flaws," ACM Computing Surveys, 26 (3), September 1994, np.

C. Carver & U. Pooch, "An Intrusion Response Taxonomy and its Role in Automatic Intrusion Response," Proc. 2000 IEEE Workshop on Information Assurance and Security, US Military Academy, West Point, MD, 6-7 June 2000, np.

Biosemiotics web page at http://everest.ento.vt.edu/~sharov/biosem (e.g. Sharov, "Signs and Values", 1997, Ogryzko, "Physics and Biosemiotics")

G. Booch, Object-oriented Analysis and Design, Addision-Wesley, 1994.

G. Booch, J. Rumbaugh, I. Jacobson, The Unified Modeling Language User Guide, Addision-Wesley, 1999.

S. Kauffman, At Home in the Universe, Oxford University Press, Oxford, UK, 1996.

National Research Council, Trust in Cyberspace, National Academy Press, Washington, DC, 1999.

Halpern, Moses, "Knowledge and Common Knowledge in a Distributed Environment," IBM Almaden Shirey, "Internet Security Glossary", RFC 2828.

Other taxonomy papers by: Denning and Branstad (1996), Bishop (1995), Aslam (1995), Shostack and Blake (1999), Halme and Bauer, Rogers (1999) also reviewed

Information Assurance – A Complex Systems Perspective

Dr. Vipin Gopal
Honeywell Technology Center

Abstract

Evolution of science from time immemorial has seen efforts by mankind to characterize, analyze, and influence what exists and happens in our environment of interest. Today, our understanding of real-world systems has reached a point that we would like to describe as, at a higher level of *complexity*. Consequently, the scientific community is making great strides towards efficient management of *complex systems*—finding solutions to previously unanswered questions, but at the same time, unraveling newer and tougher challenges. The increasingly competitive business environment is aiding this effort—a strive for better efficiencies and with faster time-to-market directives, there is an ever increasing need for addressing such challenges. One could find such intense scientific endeavor in multiple diverse disciplines, ranging from management of the

supply chain in a global enterprise to human societal networks, and immunology and drug development.

Fundamentally, there exist some basic underlying themes in the *complex systems,* that engineers and scientists are trying to gain a deeper understanding of. Complex systems of interest today have many of the following characteristics— they are large-scale (systems of systems), coupled (presence of interacting components), nonlinear (analytical intractability), hybrid (non-homogenous elements in behavior characterization), non-stationary (continuous evolution), uncertain (incomplete information), and fuzzy (qualitative and quantitative criteria). Different communities are gaining better understanding of such aspects, from vastly different angles, dictated by their domains of interest. However, an increasingly encouraging trend is beginning to emerge—with the development and evolution of complexity theory, more and more scientists are beginning to understand the fundamental parallels in problems across multiple disciplines, and subsequently, leveraging the concepts developed in sister domains. In this talk, we will discuss the features of the intrusion tolerance problem, and explore how similar features have been addressed in other domains.

As an example, we will specifically look at the *hybrid* aspect of intrusion tolerance. An intrusion tolerant system seeks to maximize integrity and availability in the enterprise to satisfy mission objectives. Simultaneous manipulation of discrete and continuous control leads to efficient and superior facilitation of availability and integrity properties, a topic that is not very well understood today. Discrete actions include access level modifications and filtering rules, and continuous actions consist of allocation and re-allocation of computing and communication resources. We draw parallels to similar problems that have been addressed in other complex systems, including hybrid control in refineries and chemical plants, and multi-model hybrid control in battlefield management.

References

V. Gopal, "Mathematical Programming Approaches for the Control and Optimization of Hybrid Systems," Proc. DARPA-JFACC Symposium on Advances in Enterprise Control, San Diego, CA, 1999, pp111-120.

D. Godbole, V. Gopal, J. Jelinek, B. Morton, D. Musliner, & T. Samad, "Multi-Model Predictive Control of Military Air Operations," Proc. DARPA-JFACC Symposium on Advances in Enterprise Control, San Diego, CA, 1999, np.

WORKING SESSION RESULTS

This section summarizes the results of the discussions from the seven working sessions of the workshop. The first working session gave the participants to respond to the formal presentations as a group. The next two working sessions considered the state of the development of theories describing information system behavior. The following two sessions addressed the issues associated with validating these theories through various forms of experiments. The final sessions developed the conclusions and recommendations from the workshop.

As one might expect from a workshop exploring such a broad topic as information system phenomena and considering the diverse backgrounds of the workshop participants, the discussions in the working sessions moved fluidly and rapidly over the space of information system descriptions. As a result, the working session topics did not constrain, as they should not, many of the interactions during those sessions. This required reorganizing much of the information produced by these sessions into the two major categories presented below. However, every effort was made to minimize the loss of any information from the working session discussions.

Theoretical Description of Information System Phenomena

Session Leaders: Michael Frentz & Vipin Gopal

This set of working sessions began with the goals of

- Broadly characterizing the coverage of the current theory describing the behavior of complex information systems, and
- Identifying specific areas where experiments are needed to support further theory advancement.

The first of the theory sessions began by reviewing the results of the DARPA IA Dark Spaces Workshop. These results defined a specific application context for the information systems of interest to DARPA, the workshop sponsor, and provided a touchstone to which the participants returned throughout the workshop when they needed to constrain their discussions to a particular application.

Within this group of sessions, the participants

- Defined the phenomena underlying information system behavior,
- Identified the elements of the existing theory base with which to describe information system behavior,
- Summarized the particular theories relevant to IA and linked those to the broader theory base, and

- Identified several outstanding challenges related to describing those information systems relevant to IA.

Phenomena of Information Systems

[We should be asking] "... what are the key parameters of information systems rather than what is the definition of information systems." Scott Evans suggested. This suggestion prompted a discussion that revolved around identifying and characterizing the macroscopic phenomena underlying complex information system behavior.

This discussion began by searching for consensus among the participants of what constituted a macroscopic phenomenon of information systems. Common usage, as represented by a general purpose dictionary, defines a phenomenon as an observable event [1]. Dr. Belyavin specified that definition for the workshop's purposes by recommending that a macroscopic phenomenon was one that impacts the information system's mission. "In fact," he observed, "we were really looking at two separate things, the macroscopic phenomena and the consequences of disturbances to the macroscopic phenomena."

Despite considerable discourse on the topic that considered several examples, the participants did not finally agree upon the definition of phenomena as related to information systems. This resulted largely due to the time constraints imposed upon the working session. Within the given time, they did identify several properties of information and information systems as well as numerous dependencies between these properties that were relevant to IA applications. These properties are organized into three groups below describing

- Information system structure,
- Information system dynamics, and
- Information itself.

Tables 1-3 present the properties describing information systems together with the informal definitions of each.

Table 1. Properties Describing Information System Structure.

Property	Description
Size	total number of individual resources associated with an information system.
Interconnectivity	number of processing nodes with which each processing node can communicate. Statistical descriptions of interconnectivity (e.g., average interconnectivity) may be necessary to characterize nonhomogeneous and dynamic systems.

Table 1. Properties Describing Information System Structure (continued).

Property	Description
Diameter	average number of links traversed to get from one place in the system to another.
Topology	interconnection pattern of nodes on a network [2].
Intelligence	ability that an information system possesses to recognize conditions and respond to them appropriately.
Complexity	degree to which a system or component has a design or implementation that is difficult to understand and verify [2].
Response Latency	time interval required for the system to respond to a stimulus. This interval may vary over time and so may require a statistical description.

Table 2. Properties Describing Information System Dynamics.

Property	Description
Growth Rate	rate at which processing, storage and communications resources are added to the system over time. This property may also be expressed in terms of the rate at which nodes and links are added to the system topology.
Stability	(1) an aspect of system behavior associated with systems having the general property that bounded input perturbations result in bounded output perturbations [2, 3]. (2) the ability of a system to continue function unchanged despite disturbing or disruptive events [2].
Robustness	degree to which a system or component can function correctly in the presence of invalid inputs or stressful environmental conditions [2]. In the context of this workshop, robustness specifically related to resistance to attack.
Availability	degree to which a system or component is operational and accessible when required for use. Robustness can be expressed as a probability or percentage of time available over the time needed [2, 4].
Integrity	degree to which a system or component prevents unauthorized access to, or modification of, its computer programs or data [2].
Agility	degree to which a system can respond to change.

Table 2. Properties Describing Information System Dynamics (continued)

Property	Description
Adaptability	degree to which a system can adjust its structure or function to suit a specified use or situation [1]. In the context of this workshop, adaptability related specifically to the ability to resist the effects of an attack or failure.
Vulnerability	degree to which a system's function or information can be degraded or compromised by attack or failure.
Recoverability	ability of an information system to recover from the effects of an attack or failure.
Usability	ease with which a user can learn to operate, prepare inputs for, and interpret the outputs of an information system or component [2].
Maintainability	ease with which a system or component can be modified to correct faults, improve performance or other attributes, or adapt to environment changes [2].

Table 3. Properties Describing Information Characteristics.

Property	Description
Velocity	rate at which information traverses the distance from one part of a system to another. Information velocity may require statistical description when applied to different types of information at different times.
Flow Rate	rate at which information moves through the components of the system. Information flow rate may require statistical description when applied to many components at different times.
False information flow rate	rate at which false information spreads within a system. As with other flow rates, this may require statistical descriptions when applied to the aggregate behavior of several components or over time.
Adverse information flow rate	rate at which adverse information (e.g., computer viruses) spreads through a system. As with other flow rates, this may require statistical descriptions when applied to the aggregate behavior of several component or over time.

Table 3. Properties Describing Information Characteristics (continued).

Property	Description
Effects of False Information	potential for false information to have an effect upon the functions of a system. This property may depend upon the false information flow rate.
Degradation	degree to which information is degraded when it flows through the system. This may be related to the loss of value of information to the system as it is manipulated.
Integrity	completeness and accuracy of information, especially after it has been manipulated in some way [4].
Insecurity Potential	potential for information to flow unintended out of the system. This property could describe such security vulnerabilities of a system as holes through which information flows.
Propagation Difficulty	amount of effort required to propagate information from one place in a system to another. This property could be seen as analogous to an impedance to information flow.

Discussion of the properties of information systems yielded several important general points.

- Dr. Lee Benzinger pointed out that many of these properties are emergent properties where an emergent property is one in addition to the functionality needed to support a specific mission. Developers may design an information system to exhibit one or more emergent properties such as availability and integrity.
- Dr. Benzinger also suggested that normality, for an information system, is a state where the system supports the mission. All deviations from normality threaten to degrade the system's ability to support its missions.
- [Availability and vulnerability] " ... tend not to be measured independently. Security only depends upon the mission. Security terms must be defined in terms of the mission." Mr. Michael Frentz
- In wrestling with defining information system phenomena, Dr. Hassen Saidi suggested that system behavior could be described by a set of variables, sequences of those variables, and interpretations of those sequences.
- [Providing system security] " ... is a balance between agility and stability." Dr. Andrew Belyavin

- "We talk about data, information (that is, fused data) and knowledge. To insure information integrity, we need some level of data integrity. To insure knowledge integrity, we must have information integrity." Dr. Andrew Belyavin
- "These lists of properties are neither complete nor independent. In some cases (e.g., vulnerability, usability), the community has not agreed upon the definition of or means to quantitatively measure these properties. In order to develop meaningful descriptions of information system state, we need to articulate the theory that specifies a reasonably complete set (for a particular group of applications like IA) of independent measurable properties." Mr. Scott Harmon

Of course, many of the properties presented in Tables 1 through 3 above are not independent of one another. The workshop participants identified several property couplings that were important to the design and use of information systems. They paid particular attention to the relationships that impact the IA characteristics of a system. Table 4 shows the property couplings that the participants identified.

As with the discussion of information system properties, the discussion of the relationships between those properties elicited several interesting and important points.

- Drs. Saidi and Cox emphasized the likely relationship between complexity and growth rate of an information system and suggested that this dependency could be readily explored experimentally since both properties were quantitative and observable.

Table 4. Possible Relationships that Exist between Information System Properties.

Property	Complexity	Content	Stability	Vulnerability	Size
Growth Rate	X				
Robustness	X				
Flow Rate		X			
Adaptability			X		
Agility				X	
Maintainability			X		
Recoverability				X	
Usability				X	
Intelligence					X

- Dr. Roger Cox illustrated the dependency between information content and flow rate with the example of counterfeit money. If the members of an economic community suspected the existence of counterfeit currency of a particular denomination within the community

then that denomination tended to flow faster within the system because people were trying to get rid of it while it still had value.

- Dr. Belyavin suggested that stability and agility might well be two ends of a spectrum of ability. An extremely stable system would resemble the tortoise while an extremely agile system would behave more like the hare. Clearly, any robust system must possess some balance between both interrelated properties.
- Dr. Belyavin also indicated that a similar balance might exist between a system's vulnerability and its ability to recover from attack or failure.
- Dr. Steven Bush found another balance between the properties of vulnerability and usability. Making a system invulnerable also makes it hard to use.
- Prof. Tommaso Toffoli discussed the work of R.W. Keyes and R. Landauer [5]. These authors found that the size of individual memory devices should be on the order of 1/e to minimize the probability of their states being affected simply by the effects of thermal noise. This result illustrates the struggle between additivity and multiplicativity, i.e., the compromise between being massive and being intelligent. This same observation shows the importance of the level of aggregation. Aggregating a system's functionality too much or too little both yield the wrong balance.
- Prof. Toffoli also suggested that a system's vulnerability could be reduced by reducing the amount of documentation associated with a system's implementation. He argued that if one could automatically generate a system implementation (particularly, software) then one need only retain and protect a single copy of the high level design documentation. This lack of intermediate documentation would force attackers to reverse engineer a system to compromise it, a difficult task for a complex system. Mr. Samuel Nicholson indicated that while this was true, the lack of sufficient documentation would also hinder maintenance and upgrades.
- Mr. Harmon felt that, again, these relationships between properties, even the ones described, were not complete enough to be very useful. Any useful information theory should consistently and quantitatively describe the dependencies of information system properties, particularly with their time dependencies if possible.
- Dr. Benzinger described how system architects create layers and hierarchies in their designs. In doing this, they trade complexity and robustness by allocating system functionality within these partitions.

Existing Theory Base Describing Information Systems

Explaining the phenomena underlying information system behavior, particularly in the context of IA, can draw from a very large base of theoretical developments. The workshop participants identified several contributors to this base.

- Complexity theory
- Control theory
- Economic theory
- Evolutionary theory
- Game theory
- Human factors
- Immunology and epidemiology
- Information theory
- Mathematical logic
- Modeling and simulation
- Network equilibrium analysis
- Physics of computation
- Probability theory
- Reliability theory
- Systems theory
- Theory of computation
- Thermodynamics & statistical mechanics

This list provides insight into the contributing technical literature collections at the highest levels of abstraction. For example, control theory includes linear control theory, optimal control theory and adaptive control theory as well as related such related areas as parameter identification, estimation and prediction.

Dr. Cox suggested that if the notion of information system behavior includes utilization then the work describing network utilization analysis is also applicable. This work uses normative modeling to derive equilibrium solutions for how users employ a network to maximize their benefits. These techniques, called network equilibrium analysis, have been applied extensively in network planning.

Dr. Cox also recognized that by generalizing the notion of information systems to include economies, the very large base of economic theory, particularly market theory, becomes available. Mr. Frentz indicated that the economic modeling community is presently applying game theory to their problems. While game theory is broadly useful in describing some information system behavior, those issues associated with IA generally require multiplayer games and that requirement adds significant complexity to the problem. In fact, some problems in multiplayer game theory cannot be solved with existing techniques.

Like control theory, systems theory includes a wide range of topics including general systems theory, hybrid systems theory, cellular automata, systems design methodologies, system design capture, brittle systems theory, composition theory and autonomous system modeling. Dr. Belyavin suggested that a general theory of IA has some common building blocks but it will mostly draw from general systems theory and process analysis. The work of Keyes and Landauer [5], discussed earlier, provides a connection to another very large literature set of the physics of computation. A bibliography of this literature is included in the final report of the project that organized this workshop [6]. Finally, Prof. Clarke, Mr. Frentz and Mr. Evans indicated that none of these theories provide sufficient machinery to explain and predict information system behavior and that this situation might require an amalgam of the current theories. Mr. Frentz asked if this was akin to finding the science behind war, another highly interdisciplinary endeavor. Mr. Evans suggested that any theory of IA must draw just the right bits from all of these theory bases and connect them meaningfully.

Theory Relevant to IA

Michael Frentz justifiably began the working session he lead by exploring the results from the DARPA Dark Spaces Workshop. This workshop identified the many technologies that contributed to IA and characterized the state of development of each of these. The technology problem areas were identified as the dark spaces of IA. This stimulated pointed discussion about IA and several illuminating observations surfaced.

- Mr. Frentz identified three types of beings in IA cyberspace: defenders, collaborators and attackers.
- "Security only depends upon the mission. Security terms must be defined in terms of the mission." Mr. Michael Frentz
- "In the context of the mission, one can define threats and once having defined threats then one can define attacks." Dr. Lee Benzinger
- "Security creates a symmetry argument. ... Whenever you get smart, your enemy also gets smart. ... This creates a competition. ... Whenever you've succeeded, you create the grounds for your own demise." Prof. Tommaso Toffoli

Insight into IA considerations and the results of the Dark Spaces Workshop enabled the participants to identify the ways in which each of the theory bases could contribute to solving the IA problems. Table 5 describes these associations.

Table 5. Contributions of Information System-Related Theory Base to IA Problems.

Theory Base	Contribution to IA Problems
Complexity theory	Modeling/Simulation, Course of Action Projection, Adaptive Survivable Network Architectures, Cryptology, Formalized Design, Dynamic Coalition
Control theory	Cyber Strategy, Modeling/Simulation, Cyber Sensor Exploitation, Situation Understanding, Autonomic Response, IA Sensors, Intrusion Detection, Formalized Design, Auto Forensics, Course of Action Projection, Adaptive Survivable Architectures, Adaptive Survivable Network Infrastructures, Protective Mechanisms, Physical Security, Security of Mobile Agents
Economic theory	Cyber Strategy, Modeling/Simulation, Intrusion Assessment, Situation Understanding, Course of Action Projection, Policy, Dynamic Policy, Dynamic Coalition
Evolutionary theory	Modeling/Simulation, Autonomic Responses, Adaptive Survivable Architectures, Adaptive Survivable Network Infrastructures, Protective Mechanisms, Dynamic Policy
Game Theory	Intrusion Assessment, Cyber Sensor Exploitation, Situation Understanding, Cyber Strategy, Modeling/Simulation, Lifecycle Attacks, Course of Action Projection, Dynamic Coalition, Auto Forensics, Autonomic Response, Formalized Design, Insider Attacks, Physical Security, Protective Mechanisms, Dynamic Policy
Human Factors	Intrusion Assessment, Modeling/Simulation, Situation Understanding, Formalized Design, Dynamic Policy, Policy, Law Enforcement Policy
Immunology and Epidemiology	Autonomic Response, Formalized Design, Security of Mobile Agents, Dynamic Policy, Dynamic Coalition

Table 5. Contributions of Information System-Related Theory Base to IA Problems (continued).

Theory Base	Contribution to IA Problems
Information theory	Situation Understanding, Modeling/Simulation, Cryptology, Security of Mobile Agents, Dynamic Coalition
Mathematical Logic	Semantic Assurance, Intrusion Detection, Cyber Sensor Exploitation, Modeling/Simulation, Intrusion Assessment, Situation Understanding, Formalized Design, Composable Trust, Auto Forensics, Multi-Domain/Multi-Level Security, Dynamic Coalition, Security of Mobile Agents
Modeling and Simulation	Cyber Strategy, Modeling/Simulation, Situation Understanding, Intrusion Assessment, Formalized Design, Auto Forensics, Course of Action Projection, Adaptive Survivable Architectures, Protective Mechanisms, Dynamic Coalition, Insider Attacks, Security of Mobile Agents
Network equilibrium analysis	IA Sensors, Modeling/Simulation, Cyber Strategy, Adaptive Survivable Network Infrastructures, Formalized Design, Policy, Dynamic Policy, Dynamic Coalition, Adaptive Survivable Architectures, Security of Mobile Agents
Physics of computation	Cyber Strategy, Modeling/Simulation, Formalized Design, Course of Action Projection, Security of Mobile Agents, Dynamic Coalition
Probability theory	Intrusion Assessment, Cyber Sensor Exploitation, Situation Understanding, Cyber Strategy, Modeling/Simulation, Auto Forensics, Intrusion Detection, Composable Trust, Lifecycle Attacks, Formalized Design, Insider Attacks, Physical Security, Adaptive Survivable Network Infrastructures, Course of Action Projection, Adaptive Survivable Architectures, Security of Mobile Agents, Multi-Domain/Multi-Level Security

Table 5. Contributions of Information System-Related Theory Base to IA Problems (continued).

Theory Base	Contribution to IA Problems
Reliability theory	Cyber Strategy, Modeling/Simulation, Situation Understanding, Auto Forensics, Formalized Design, Course of Action Projection, Adaptive Survivable Network Infrastructures, Adaptive Survivable Architectures, Security of Mobile Agents
Systems Theory	Situation Understanding, Modeling/Simulation, Formalized Design, Course of Action Projection, Dynamic Policy, Dynamic Coalitions, Insider Attacks, Adaptive Survivable Network Infrastructures, Adaptive Survivable Architectures, Protective Mechanisms, Dynamic Coalition, Security of Mobile Agents
Theory of computation	Semantic Assurance, Cyber Strategy, Modeling/Simulation, Intrusion Assessment, Formalized Design, Composable Trust, Course of Action Projection, Protective Mechanisms, Dynamic Coalition, Security of Mobile Agents
Thermodynamics & statistical mechanics	Cyber Sensor Exploitation, Modeling/Simulation, Formalized Design, Adaptive Survivable Architectures

The process of identifying the theory base that can contribute to describing the macroscopic phenomena of complex information systems evoked several relevant comments from the participants.

- Dr. Bush noted that the concepts behind immunology and epidemiology have not been broadly applied to IA and certainly seem to be solving analogous problems.
- Dr. Belyavin felt that if one can do good situation understanding then one could also do good course of action projection.
- Prof. Toffoli is putting a plan together to decrease the vulnerability to certain types of attacks at design time. This approach is based upon the observation that the need to reverse engineer a complex system provides strong protection. He suggests protecting the master design document with existing cryptographic techniques that make detection of tampering and access possible then destroying all the records of the intermediate

development (e.g., code design documents). He posits that one is only open to attack when the master design is left unprotected. His approach will automatically turn high level specifications into a final working product. This approach fits into protective mechanisms.

- Dr. Saidi asked whether the technology underlying human interface design (e.g., human factors) contributes to developing good models of people. To this Dr. Belyavin replied that the human factors community has theory but that it does not provide very good predictions of real human behavior.
- Dr. Benzinger noted that the elements of composition could help in policy, dynamic policy, dynamic coalitions, multi-level domain/multi-level security and insider attacks. New components may be needed for coalitions and multi-level domain/multi-level security and composition theory could help in defining those.
- Dr. Gopal observed that course of action projection requires building structures describing the command hierarchy. This work draws from game theory and general systems theory. His group has done some experimental work in this area and one could adapt the knowledge gained from these experiments to course of action projection.
- Nearly all of the participants agreed that the theory base addressing information system phenomena can contribute to IA design guidance and system vulnerability analysis. These two areas, in turn, apply broadly to most of the problem domains defined by the Dark Spaces Workshop.

Most of the participants were exploring specific theories of their own that represented either specializations of the broad areas described in Table 5 or combinations of those areas. Table 6 shows the correspondence that currently exists between the specific theoretical approaches of the participants and the general theory bases.

Table 6. Correspondence between Individual Participant Theories and Major Theoretical Bases.

Participant Theory	Corresponding Theoretical Bases	Participant
Automated Design	Systems Theory, Mathematical Logic, Probability Theory	Toffoli
Category Theory	Mathematical Logic, Modeling & Simulation	Clarke
Composite Systems	Modeling & Simulation	Belyavin
Cyberlogic	Mathematical Logic	Saidi

Table 6. Correspondence between Individual Participant Theories and Major Theoretical Bases (continued).

Participant Theory	Corresponding Theoretical Bases	Participant
Cyberspace Cartography	Human Factors, Mathematical Logic	Frentz
EPI	Mathematical Logic & Probability Theory	Cox
Hybrid Systems	Complexity Theory, Systems Theory, Mathematical Logic	Gopal
IAM	System Theory, Mathematical Logic	Benzinger
Information Physics	System Theory, Theory of Computation, Physics of Computation, Thermodynamics & Statistical Mechanics, Information Theory, Complexity Theory	Harmon
Kolmogorov Complexity	System Theory, Theory of Computation, Information Theory	Bush & Evans

Outstanding IA Theory Challenges

Despite the considerable relevant theoretical base and the efforts of the individual researchers, the workshop participants identified several challenges that remain in describing information system behavior, in general, and IA, in particular. Table 7 summarizes these challenges together with relevant comments made by the participants.

Table 7. Summary of Outstanding IA Challenges.

Challenge	Relevant Comments
Vulnerability Analysis	"Vulnerability analysis is key to most of these [the IA dark spaces] bubbles. It's a fundamental problem. Identifying the vulnerabilities is the primary challenge not covered by existing theory. We need some way to quantify the size of security holes." Dr. Steven Bush
	[Vulnerability analysis] " ... is pretty much black magic now." Dr. Lee Benzinger
	"We can't design systems that are 100% secure and need an assessment methodology that can determine vulnerability. We then must be able to decompose this requirement into different parts." Dr. Hassen Saidi
IA Design Guidance	"The ways we build systems will impact many of these [Dark Spaces] bubbles. We're not given any information on how to design systems. Some guidance to designers to gain these [IA] properties is needed." Mr. David Gross
People Models	"Situation understanding is where models of people might be most useful. We're not very good at describing what people contribute to situations." Dr. Andrew Belyavin
Requirements Translation	"Refining high level requirements into detailed specifications is not well understood." Dr. Hassen Saidi
Event Detection & Measurement	[We need to know] "... what to sense to assess vulnerability. We're already making sensors without agreement of what the fundamental parameters are. We must develop laws from which to derive metrics that could then be used to guide [IA] sensor development." Mr. Scott Evans
Information Value	"Another important issue is the value of information, to the people that own it but also to attackers. This includes corrupting the information for the owner and compromising the information itself. Value is different for different people. If two attackers team then the information may have enough value to the two to make it cost effective to mount an attack. The most effective attack might be to create a requirement that is so costly that the system will never be built." Mr. David Gross

Table 7. Summary of Outstanding IA Challenges (continued).

Challenge	Relevant Comments
Situation Assessment	"Situation assessment and course of action projection are tightly coupled. Situation assessment lies at the heart of the information system problem and it's often skated over." Dr. Andrew Belyavin
	"One area where there's a gap is between detection and response, called correlation." Dr. Lee Benzinger
System Recovery	"Almost no work has been done in the area of recovery." Dr. Lee Benzinger
Mapping Microscopic to Macroscopic Phenomena	"We don't understand the inner level mappings [of an information system], what microscopic variables have effects upon the macroscopic properties. The dependencies depend upon the intermediate decomposition. Therefore, the system architecture should have an impact upon the relationships. Are there heuristics for developing a system architecture that lead to robust systems?" Dr. Lee Benzinger
IA Costs	"We don't have a handle on the costs of implementing IA." Dr. Steven Bush
	"Without the costs we can't do design trades." Dr. Lee Benzinger
Design to Behavior Dependencies	"We need to understand the overall macroscopic behavior of the system. We don't understand how changes in the system design cause the overall behavior of the system to change." Dr. Vipin Gopal
Intelligent Attackers	Prof. Thomas Clarke pointed out that an intelligent attacker wasn't considered by Shannon. This is a major shortcoming of present day information theory.
Vulnerability & Availability	"These things [availability and vulnerability] tend not to be measured." Mr. Michael Frentz
	"I have yet to see any rigorous analysis of availability and vulnerability." Dr. Vipin Gopal
IA Framework	"We may need a[n IA] framework because theory is not sufficient." Dr. Lee Benzinger

Table 7. Summary of Outstanding IA Challenges (continued).

Challenge	Relevant Comments
Understanding Teams & Teamwork	"We have a limited understanding of how teams work and what the effects of individuals are upon the behavior of teams." Dr. Andrew Belyavin
Threat Models	"A good theory could give us abstractions of a threat against which we could test design decisions." Dr. Lee Benzinger
Interdependency & Robustness Measures	"We need a measure of interdependency or robustness." Mr. Scott Evans

Experimental Validation of Information System Theories

Session Leaders: Andrew Belyavin & David Gross

This set of working sessions began with the goals of

- Identifying specific opportunities for collecting experimental data that would support characterizing information system behavior,
- Broadly characterizing the nature of the existing body of experimental data, and
- Identifying the experimental opportunities associated with validating specific theories describing information system phenomena.

Experimental validation of any information system theory can take one of three forms:

- Controlled formal experiments
- Observation of existing system behavior
- Simulation of system behavior

In controlled formal experiments, the information system under observation and the environment within which it operates are carefully controlled; system behavior is observed through instrumentation with well understood error characteristics; and the system is operated under very constrained and well characterized conditions. These experiments should lead to data that describes the system's behavior in well understood situations. Observations of existing system behavior have the experimenter playing a strictly passive role. In this form, the experimenter can control almost nothing, if anything at all, about the system or the surrounding environment. Instead, the researcher must carefully characterize the instrumentation and measurement processes to understand the influence these aspects have upon the resulting data. In addition, the experimenter must observe all conditions that may influence system behavior in order to understand the magnitude of their contributions. In some cases, experimenters can mine data on information system behavior from existing databases. Where existing data is used then the experimenter must endeavor to understand the conditions under which that data was collected and the measurement apparatus used to collect the data. Statistical techniques may alleviate the need to completely understand the nature of the prevailing conditions but they bring their own limitations. Finally, some experiments can be conducted using simulations of real systems and the surrounding conditions. Simulations can often make experiments significantly easier and more practical to run but they introduce the effects of the abstraction inherent to all simulations (i.e., they do not really represent the real world but some subset of it). The primary concern in using simulation is the validity of that representation (i.e., its correspondence with what actually occurs). Obviously, each of these forms of experimental validation has different advantages and disadvantages.

The working sessions contributing to this section produced

- An annotated list of sources of existing data describing the behavior of various information systems,
- An annotated list of resources that could support simulations of information systems at various levels,
- A list of experimental facilities within which controlled experiments on or observations of information systems could be conducted,
- A list of challenges facing experiments exploring the behavior of information systems, and
- A collection of guidance the workshop participants suggested when considering conducting information system experiments.

In addition, the participants identified how they could exploit the different experimental opportunities to validate their own theories describing information system phenomena.

Existing Data Sources

Mining existing sources of data about information system behavior presents one approach to validating theories describing information system phenomena. The workshop participants identified four possible general sources of this data:

- Existing databases
- Organizations with data
- Published bodies of experimental data

In using any existing data sources, especially those with anecdotal information, Dr. Cox cautioned, "We need to be concerned about urban myths."

Existing Databases

The workshop participants identified six different sources of existing data:

- Economic Databases
- DARPA IA Red Team Experiment Results
- Mitre Common Vulnerabilities and Exposures Resource
- Bugtraq Discussion List Archives
- MIT Lincoln Laboratories Intrusion Experiment Database
- Synthetic Theater of War Experiment Databases

The paragraphs below discuss each of these sources to some degree.

Economic Databases. Dr. Cox noted that some evidence exists to support the success of some economic theories in predicting system behavior. Economists try to predict the effects of various macroscopic phenomena (e.g. money supply, inflation rates) consistent with the psychology, sociology and technology associated with microscale behaviors (i.e. microeconomics). To do this, they observe the treatment of various financial instruments (e.g., currency, equities, debt), all carriers of information. Such components of economics as macroeconomics and international finance are concerned with characterizing the macroscopic phenomena of complex information systems in a very real way. Several economics databases exist that hold volumes of experimental data from which to draw. Also, many economists have analyzed this data to support their theories and published the results of these analyses. These analyses may offer further data to support validating theories of more general information system behavior.

DARPA IA Red Team Experiment Results. Dr. Benzinger and Mr. Frentz recommended the data collected from the IA Red Team experiments. In these experiments, a group of individuals playing the role of attackers (i.e., the Red Team) worked against another group playing the role of defenders (i.e., the Blue Team) in various IA scenarios. Experimenters collected extensive data in these exercises. These data include such information as attack trees, attacker-defender interactions and attack timelines. Currently, the records of these experiments reside behind password protection and access must be granted through DARPA. The need to limit access to these data prevents publishing a link to them in this document.

Mitre Common Vulnerabilities and Exposures (CVE) Resource. CVE is a dictionary of standardized names for vulnerabilities and other information security exposures. CVE aims to standardize the names for all publicly known vulnerabilities and security exposures. Through this it wants to makes sharing data across separate vulnerability databases and security tools easier. CVE contains pointers to other relevant databases and through these may make it easier to search for information in these databases. However, CVE should not be considered as a vulnerability database on its own merit. The MITRE Corporation maintains CVE and moderates Editorial Board discussions. The current version of CVE is free to use and available for download from this Web site: www.cve.mitre.org.

Bugtraq Discussion List Archives. Bugtraq is a widely circulated mailing list where subscribers detail the faults and flaws they discover in commercial software products. The Bugtraq list archives constitute a very large database of this information and create another resource related to IA from which information system behavior data may be mined. The Bugtraq mailing list can be accessed through mailman.newdream.net/mailman/listinfo/bugtraq.

MIT Lincoln Laboratories Intrusion Experiment Database. Dr. Benzinger noted that MIT Lincoln Laboratories has been testing various intrusion detection systems against a set of standardized attack scenarios for many years and have collected these data into a database of the results. Again, these results are maintained behind password protection and only DARPA and Lincoln Laboratories can grant access to this database.

Synthetic Theater of War (STOW) Experiment Databases. STOW is a DOD Advanced Concept Technology Demonstration (ACTD) that is developing, integrating and transitioning the technologies necessary to demonstrate the potential of high resolution (platform level) simulation to support joint command and staff training and mission rehearsal. The operational sponsor is the United States Atlantic Command (USACOM) and the technology developer is the Defense Advanced Research Project Agency (DARPA). STOW has run many experiments and collected substantial data from these activities. These data illustrate how battlefield entities exchange information under a variety of simulated conditions. The STOW website is at www.darpa.mil/iso/stow/. This site provides general information about STOW. However, one can probably only access the databases from STOW experiments with permission from the operational sponsor and technology developer.

Dr. Benzinger observed that a lot of these data are low level. It is the microscopic data from which macroscopic information could be inferred.

This list of databases relevant to information system phenomena is not complete but rather represents examples that an experimenter might exploit.

Organizations with Data

The workshop participants identified several organizations that maintain databases with information relevant to characterizing information system phenomena:

- National Communications System National Coordinating Center for Telecommunication
- Cooperative Association for Internet Data Analysis
- Software Engineering Institute CERT Coordination Center
- Alliance for Telecommunications Industry Solutions
- North American Energy Reliability Council
- Federal Aviation Administration Air Traffic Services
- Centers for Disease Control and Prevention

The paragraphs below provide some information on these organizations and, where possible, on the relevant databases they maintain.

National Communications System (NCS) National Coordinating Center for Telecommunications (NCC). The NCC operates under the management of the NCS which, in turn, operates as part of the Defense Information Systems Agency (DISA). The NCC has established an Information Sharing and Analysis Center (ISAC) function as part of its national security/emergency preparedness (NS/EP) telecommunications mission. Under the ISAC concept, the NCC gathers, analyzes, and disseminates private sector and Government information to participating telecommunications entities. The NCC is thus a central hub for sharing critical NS/EP telecommunications information on vulnerabilities, threats, intrusions, and anomalies among

participating companies and between industry and the Government. Actually, two ISACs were established at the same time, the Telecommunications ISAC at the NCC, and the Banking and Finance ISAC. Both of these centers have databases that may provide information relevant to understanding information system phenomena. The NCC's website is at www.ncs.gov/ncc/.

Cooperative Association for Internet Data Analysis (CAIDA). CAIDA collects data on the internet that could support theory validation. CAIDA is a collaborative of organizations in the commercial, government, and research sectors that promotes cooperation in the engineering and maintenance of a robust, scalable global Internet infrastructure. CAIDA provides a neutral framework to support cooperative technical endeavors. Its goals include (1) encouraging the creation of Internet traffic metrics (in collaboration with IETF/IPPM and other organizations); and working with industry, consumer, regulatory, and other representatives to assure their utility and universal acceptance; (2) creating a collaborative research and analytic environment in which various forms of traffic data can be acquired, analyzed, and (as appropriate) shared; and (3) fostering the development of advanced methodologies and techniques for traffic performance and flow characterization, simulation, analysis, and visualization. CAIDA resides at the San Diego Supercomputing Center (SDSC), an extension of the University of California at San Diego (UCSD). Its website URL is www.caida.org

Software Engineering Institute (SEI) CERT Coordination Center (CERT/CC). The CERT Coordination Center is part of the Networked Systems Survivability Program at the Software Engineering Institute at Carnegie Mellon University. It studies Internet security vulnerabilities, provides incident response services to sites that have been the victims of attack, publishes a variety of security alerts, researches security and survivability in wide-area-networked computing, and develops information to help improve website security. CERT/CC has built several databases of security vulnerabilities that they use in their activities. Dr. Benzinger suggested that Mr. John McHugh is a theorist at CERT who might be a useful resource for well focussed questions on security vulnerabilities. CERT/CC's URL is www.cert.org.

Alliance for Telecommunications Industry Solutions (ATIS). ATIS is a membership organization that provides the tools necessary for the telecommunications industry to identify standards, guidelines and operating procedures that make the interoperability of existing and emerging telecommunications products and services possible. ATIS-sponsored committees and forums publish a variety of technical reports, analyses and documents of interest to telecommunications industry companies. For example, the Network Reliability Steering Committee (NRSC) works closely with the FCC in analyzing reports of major telecommunications outages. Its annual report demonstrates how the industry is ensuring that public wireline networks are operating in a reliable manner. ATIS manages a database of telecommunications incidents, primarily accidental, that categorizes the root cause of failures. ATIS has a website at www.atis.org.

North American Energy Reliability Council (NERC). NERC is a nonprofit organization formed by the electric utilities to promote the reliability of the electricity supply for North America. The mission of the new NAERO (North American Electric Reliability Organization), NERC's

successor organization, will be to develop, promote, and enforce standards for a reliable North American bulk electric system. NERC maintains several databases: Generating Availability Data System (GADS), Electricity Supply and Demand Database (ES&D), and the System Disturbance Database. While these databases do not deal with information specifically, they do characterize the behavior of a large scale information system, the continent's electricity supply grid. The NERC website is at www.nerc.com.

Federal Aviation Administration (FAA) Air Traffic Services (ATS). Air Traffic (AAT), a component of ATS, is comprised of the following offices: Tactical Operations, Planning and Procedures, Airspace Management, and Resource Management. AAT facilities consist of : Air Traffic Control System Command Center (ATCSCC), 61 Automated Flight Service Stations (AFSS), 15 Flight Service Stations (FSS), 14 Alaskan Rotational Flight Service Stations, 21 Air Route Traffic Control Centers (ARTCC), 352 Airport Traffic Control Towers (ATCT), 185 Terminal Radar Approach Control (TRACON) facilities, 2 Radar Approach Control (RAPCON) facilities, and 3 Combined Center/Radar Approach Control (CERAP) facilities. AAT manages civil and military air traffic in the navigable airspace by developing and recommending national policies and establishing national programs, regulations, standards, and procedures for management of the airspace, operation of air navigation and communications systems and facilities, separation and control of , and flight assistance to air traffic. They do this by providing for the security control of air traffic to meet the national defense requirements, developing and coordinating U.S. policies, standards, and procedures related to international Air Traffic, operating the FAA national and international flight information and cartographic programs, and exercising operational control and technical direction of the air traffic control system and line authority to the day-to-day operations of the system. In the course of these operations, AAT has collected significant data on air traffic operations that may be of use in validating theories describing complex information system phenomena. AAT's website is at www.faa.gov/ats/at/.

Centers for Disease Control and Prevention (CDC). The CDC's mission is to promote health and quality of life by preventing and controlling disease, injury, and disability. It includes the National Center for Chronic Disease Prevention and Health Promotion, the National Center for Environmental Health, the National Center for Health Statistics, the National Center for HIV, STD, and TB Prevention, the National Center for Infectious Diseases, and the National Center for Injury Prevention and Control in addition to several individual programs. The CDC maintains numerous databases on health information under the National Electronic Disease Surveillance System. This system includes data on disease transmission and treatment, hazardous substance release and effects, assisted reproductive technology success, behavioral risk factors, birth defects, cancer, HIV/AIDS, pregnancy risks, tuberculosis surveillance, and youth risk behavior. These databases can be accessed through www.cdc.gov/scientific.htm. The primary relevance of these data to information systems is through the analogy between IA and epidemiology.

Like the databases enumerated in the previous section, the organizations maintaining databases relevant to describing the macroscopic behavior of complex information systems go beyond those described above. For example, Dr. Cox suggested that the National Aeronautics and Space Administration (NASA) may have some data on the operation of the air traffic control system and Mr. Gross suggested looking at the information produced by the stock market. While both of these suggestions have considerable merit, they were not specific enough to find more direct sources of information. They should be considered in broader searches for information relevant to information system phenomena. Undoubtedly, many additional organizations exist that could contribute data for experiments exploring the macroscopic phenomena of complex information systems.

Published Bodies of Experimental Data

The workshop participants suggested two different published bodies of experimental work that may provide information about information system phenomena.

Prof. Toffoli indicated that considerable experimental work is being done in the area of quantum computation and quantum cryptography. Boston University is even exploring quantum cryptography assurance. The information physics bibliography in Reference [6] contains some references to the work in this area.

Dr. Cox indicated that the IA experiments, both past and present, have produced considerable data about information system behavior in the security context. He said that these results could possibly be used to validate components of information system theories.

Simulation Resources

The workshop participants recognized that simulation could serve as a useful tool for exploring the validity of theories describing information system phenomena. They suggested several possible simulation tools for this purpose.

- Opnet
- Extend
- Swarm
- Cellular automata tools
- Distributed simulation

While simulation promises a powerful tool for many purposes, its utility is limited by the validity of its representation. Dr. Belyavin captured this concern with the comment "In theory, we're talking about validation under operational conditions and a great difference exists between simulations and actual warfare." In order to use simulation to validate any theory, the simulation itself must represent the situations covered by the theory in a valid way. Blindly using simulation to validate a theory could easily result in incorrect conclusions.

OPNET. OPNET is a commercially available toolset for the designing and analyzing advanced communications network technologies, network devices, and protocols. The OPNET Modeler's object-oriented modeling approach and graphical editors mirror the structure and dynamics of actual networks and network components. The OPNET Modeler is used by the world's leading network technology companies and service providers. More information about OPNET can be found at www.mil3.com. Dr. Bush has used OPNET and cautioned that OPNET/BONES requires well understood distributions and may be more appropriate to support simulations of microscopic phenomena. For this reason, they have chosen to use Swarm for their experiments.

Extend. Extend is a dynamic, iconic simulation environment with a built-in development system for extensibility. It enables simulation of discrete event, continuous, and combined discrete event/continuous processes and systems. Extend's libraries of pre-built blocks support building simulations of a wide range of phenomena with a minimum of programming effort. However, users can add to these libraries to accommodate their unique situations. Extend provides a complete authoring environment and development system in a single tool. Extend does extend beyond a single tool with a family of tools tailored to serve particular industries. The core Extend tool supports engineering, operations research, scientific, and general purpose modeling applications. More information about Extend can be found at www.imaginethatinc.com.

Swarm. Swarm is a general purpose simulation package for investigating concurrent distributed systems, systems in which hundreds or thousands of autonomous agents interact with one another and with a dynamically changing environment. Swarm provides general purpose utilities for designing, implementing, running and analyzing such multi-agent systems. Swarm is a freely available package under the Library GNU Public License (LGPL) and is attainable from the Santa Fe Institute website. More information about Swarm can be found at www.swarm.org. Dr. Steven Bush and Mr. Scott Evans are using Swarm in their experimental studies of information system behavior.

Cellular Automata (CA) Tools. CA are discrete dynamical systems whose behavior is completely specified in terms of a local relation. A cellular automaton can be thought of as a stylized universe. Space is represented by a uniform grid with each cell containing a few bits of data. Time advances in discrete steps and the laws of the automaton's universe are expressed in such a form as a small look-up table. At each time step each cell computes its new state from that of its close neighbors and the look-up table. Thus, the system's laws are local and uniform. CA are computer simulations that try to emulate the way the laws of nature are supposed to work in nature. CA have been used to simulate such phenomena as chemical reactions, diffusion, wild fire progression, viscoelastic flow, dissipative structure formation, ecosystems, biological systems, thermomechanical processes, phase transition, immune systems and economic systems. Several CA tools exist including

- Cellular Automata Simulation System at www.cs.radford.edu/~dana/ca/cellular.html
- JCASim @ www.jweimar.de/jcasim/

- DDLab @ www.santafe.edu/~wuensche/ddlab.html
- Scarlet @ www.informatik.uni-giessen.de/scarlet/

Prof. Toffoli is currently using CA tools at Boston University to simulate an information system where the nodes are close to logical gates.

Distributed Simulation. Distributed simulation is a very broad term referring to simulation systems consisting of multiple individual simulations interacting through one or more communications media. Recent advances in protocols such as the Distributed Interactive Simulation (DIS) protocol group and simulation infrastructures such as the DoD's High Level Architecture (HLA) simplify many of the problems in constructing and integrating distributed simulation systems. The notion of distributed simulations theoretically permit one to collect a set of simulations that individually represent the environment and entities needed to achieve some purpose thereby minimizing the development time needed to create simulations of complex phenomena. Considerable effort is being invested in implementing distributed simulations to support training, analysis and acquisition. Assuming that the simulation components existed to represent complex information systems, they could be integrated and operated as distributed simulations to validate theories describing information system phenomena. However, despite the considerable interest in distributed simulation, significant technical questions remain open regarding the validity of simulation systems constructed from components even though the components themselves may be individually valid for the purpose in mind.

Experimental Facilities

"At first it might be useful to simulate. But, then we need to instrument a testbed." Mr. Scott Evans

Mr. Evans' quote captures the sentiments of most of the workshop participants regarding the key role that controlled experiments must play in the development of the science of information systems. The participants discussed two primary candidates to host information system experiments:

- DARPA's Technology Integration Center and
- the Internet.

These facilities will be discussed only briefly below since many much more informative sources exist that more fairly capture their capabilities and potentials.

DARPA Technology Integration Center (TIC)

The DARPA TIC provides a unique facility that could support fundamental experiments exploring information system phenomena. The Information Assurance Laboratory part of the TIC would most logically provide this experimental support. The IA Laboratory formally opened in February 1999 and has supported numerous experiments and demonstrations since

then. The IA Laboratory currently occupies three rooms of the TIC. Tables 8 and 9 summarize the capabilities that the IA Laboratory houses [7].

Table 8. Common Capabilities of the Support Infrastructure for the IA Laboratory.

Capability	Description
WAN Connectivity	Fractional T-3 (12Mbps) direct Internet Access, and 2-1/2 Class C routable addresses
Cabling	CAT-5 supporting 100Mbps Ethernet throughout
Routers	Two Cisco 3640 (IOS 12.1(5)) quad Ethernet, one Cisco 2514 (IOS 11.1(29)) dual Ethernet
Switches	Three Cisco 2924XL Ethernet switches, numerous 10/100 Mbps hubs
Firewalls	Sidewinder 5.0, Gauntlet 5.5
Printers/Plotters	HP Laserjet 8000N, HP Laserjet 4500N (color), HP DesignJet 3500CP (color plotter)
Backup	Legatto backup SW, robotic tape drive, two standalone 4mm tape drives, CD R/W Drive
OS	Solaris 2.5.1, 2.6, 7.0; 8.0; NT 4.0, Win 2000; Red Hat Linux (5.2, 6.0), BSDI, FreeBSD
Projection	50" TV (main lab, portable), Proxima High Resolution Projector (conf. room)

In addition to the hardware and software assets that are required to support the IA Laboratory functions, a number of services are provided for flexible operations. These include DHCP, DNS, NT Domain (printing, file sharing), and backup services.

Table 9. Combined Computing Capabilities Housed by the IA Laboratory.

Computing Platform	Platform Capabilities	Quantity
Pentium II/II	500MHz, 256 MB RAM, CDROM/Floppy/ 17" or 21" monitor	34
Sun Ultra 5	333 MHz, 256 MB RAM, CDROM/Floppy/ 17" or 21" monitor	24
Sun Ultra 10	360 MHz, 256 MB RAM, CDROM/Floppy/ 21" monitor	4
Gateway Pentium Laptop	300 MHz, 256 MB RAM, CDROM/Floppy	2

In addition to the on-site capabilities, the IA Laboratory configuration allows Virtual Private Network (VPN) connections directly between the facility and participant contractor and Government sites. This direct, secure, high-speed connectivity enables such activities as remote experimentation set-up and check-out and wider deployment of assets for experimentation and

demonstrations (actually using facilities at the remote end of the VPN). Since inception, this service has grown from five initial sites to 12 current sites.

The IA Laboratory supports both unclassified and classified operations. One of the IA Laboratory's rooms supports SECRET operations and an additional room is being prepared for TS/SCI operations in the future. These capabilities enable interactions with US military command and control, and intelligence assets.

With regards to using the TIC, Dr. Cox suggested that one should probably study simpler information systems before conducting experiments with the facilities of the IA Laboratory.

Internet

The Internet represents a vast and rapidly growing and evolving collection of information systems. Several organizations currently collect information on various aspects of the Internet. Parts of the Internet could even be controlled in very limited ways in order to perform controlled experiments. However, the largest amount of data about the behavior of the Internet must come from observations. Dr. Cox felt strongly that data about the Internet could contribute in the first phases of validating his theory of information system phenomena. He would be particularly interested in characterizing how people currently use the Internet.

Experimental Challenges

The workshop participants identified several challenges associated with experimentally validating information system theories, particularly those applicable to IA. These included

- Representing defense networks
- Representing system dynamics
- Autonomy & synthetic participants
- Efficient data mining tools
- Meaningful data collection
- Poor control of existing systems
- Limitations of extrapolation
- Inferring system behavior from component behavior
- Proprietary data restrictions

Representing Defense Networks. From DARPA's current perspective, the practical motivation for formulating and validating any theories explaining information system phenomena arises from the need to accurately characterize and predict the behavior of defense networks. These represent some of the most challenging information system problems due to their vast geographic extent, large diversity, immense complexity, sensitivity to errors, and rapid and wide ranged

dynamics. Further, defense networks may not lend themselves easily to large scale controlled experiments. As a result, interested researchers must rely upon data collected on defense network performance, small scale recreations of defense networks, and simulation. The distinct limitations of each of these options may force seeking hybrid solutions.

Representing System Dynamics. As Prof. Toffoli and others noted, substantive challenges in providing IA result from the natural dynamics between attackers and defenders. Defenders evolve their defensive measures when they detect and characterize their attackers. Yet, this adaptation causes attackers to evolve their own offensive strategies to maintain their chances of successfully compromising their targets. This circular action-reaction pattern creates complex system dynamics that information system theorists must not only explain but also must devise experiments to represent in order to validate their theories. These dynamics are inherent to the IA problem so any experiments must encompass this behavior. The need to represent these dynamics complicates any experiment, observations or simulations significantly since these dynamics occur at many levels of system aggregation.

Autonomy and Synthetic Participants. The DARPA IA experiments have gained considerable experience in planning and executing experiments that involve human participants both as attackers (i.e., Red Teams) and defenders (i.e., Blue Teams). These experiments have been complex, expensive, and difficult to control, characterize and interpret. These complications suggest the need for synthetic participants and autonomy in some information system experiments, both as attackers and defenders. Fortunately, some previous work exists in this area from which information system experimenters may draw. Mr. Frentz recommended that before pursuing this area, one should be cognizant of this existing work. For example, the DARPA Autonomic Information Assurance (AIA) Program is emphasizing autonomy. Further, substantial autonomy has been developed for intrusion detection systems. Drs. Saidi and Benzinger have contributed to this area. In general, the IA researchers are moving towards more automation and that may make experimentation somewhat easier by eliminating the complexities introduced by human decisionmakers. In addition, considerable research is ongoing to develop high fidelity human behavior representations. While this field is still in its infancy, despite many years of effort, it may provide some technology useful for information system experiments. However, whenever considering using automated representations of attackers and defenders, one must always be concerned about the validity of these representations. They are simulations and the validity of any simulations used in experiments will affect the validity of those experiments. Simulations of human behavior are notoriously difficult to validate.

Efficient Data Mining Tools. The workshop participants identified several databases that could provide information about the behavior of several different information systems. Most of these databases were developed for purposes other than characterizing the information systems they represent. This implies the need for data mining tools to efficiently extract and analyze the data needed to validate any theories of information system phenomena. While many general data mining tools exist and some are currently commercially available or in the public domain, all have

very steep learning curves and must be used carefully to avoid creating artifacts in the data sets they collect or in the analysis results they produce. Some attention must be focused upon developing new tools or adapting existing tools specifically for the purposes of validating theories explaining information system behavior. Certainly, data mining tools will be instrumental to any researcher seeking to characterize information system phenomena but the capabilities, limitations and peculiarities of those tools must be completely understood to avoid coming to incorrect conclusions from the data they mine.

Meaningful Data Collection. Dr. Benzinger indicated that the IA community has had trouble collecting meaningful data about the performance of their systems. They tend to collect what they can rather than what they need. This problem arises because of the lack of good models characterizing complex information systems. Better models will improve the quality of data collected and better data on information system behavior will, in turn, improve the quality of the models describing that behavior. Regrettably, the existing difficulties in collecting meaningful data may reduce the value of the data existing in databases about information systems. Initial experiments that use this data may need to carefully characterize its quality (e.g., using statistical tests, identifying and eliminating anomalies, applying normalization techniques) to ensure the accuracy of any conclusions drawn from it.

Poor Control of Existing Systems. Mr. Gross suggested that conducting controlled experiments upon existing information systems brings its own challenges because the services provided to that system's users cannot be disrupted by the experiments. This situation largely relegates experimenters to an observational role in most cases without the ability to change the experimental conditions. In general, the only option is collect data only under certain well known conditions and stop collecting data when the system behavior moves outside the acceptable windows. This requires experimenters to carefully identify the windows within which they will collect data. Such decisions may require more extensive models of information system phenomena than many current theories provide.

Limitations of Extrapolation. Dr. Cox emphasized the difficulty and danger of extrapolating from observations of actual conditions. Such extrapolations require experimenters to characterize the state spaces of information system behavior, at least those dimensions that include both the observations and their extrapolations. Interpolating and extrapolating techniques always make restrictive assumptions about the behavior of the surfaces in those state spaces. The simplest techniques assume those surfaces to be flat or linear. Models describing information system behavior statistically make similar assumptions (e.g., normal distributions, stationary distributions). Extrapolating without understanding the nature of the spaces over which the extrapolations are made will certainly lead to indefensible, and quite possibly incorrect, conclusions. However, gaining this understanding can present challenges of its own to experimenters and will certainly complicate any experiment employing existing observations.

Inferring System Behavior from Component Behavior. A problem related to extrapolating from existing observations is that of inferring system behavior from observations of component

behavior. Dr. Belyavin acknowledged that one could perform experiments upon the components of information systems and infer the behavior of systems incorporating those components from those observations. He suggested that much of the existing data describing system behavior really describes the performance of its components under very specific conditions. This presents two primary challenges to experimenters:

- Characterizing the conditions under which data reflects the effects of many interacting components (i.e., actual system behavior) and
- Identifying those conditions when system behavior is dominated by the influence of just a few (maybe only one) components.

These challenges, together with those posed by the need to extrapolate or interpolate from existing data, emphasize the need to adequately characterize the conditions under which data was collected and the behavior of the data itself. In some cases, insufficient information may be available to do this with existing data. Experimenters must therefore approach the data in existing databases cautiously. In some cases, one can improve the value of existing data by performing limited and carefully controlled experiments simply to characterize the nature of the behavior space or the magnitude of the influence of dominant components.

Proprietary Data Restrictions. Some of the most important data to characterize information system phenomena may be proprietary and, thus, unavailable to experimenters. This observation arose when Prof. Toffoli suggested that cost-benefit analysis of information system behavior could provide important and practical insight. He cited the example of commercial software firms producing products with known bugs but their public release still provided those firms benefit. He recommended that we should look at what other people do and determine the reasons for their choices. Dr. Benzinger indicated that commercial firms, such as NAI, treat cost-benefit and design tradeoff data as very sensitive information. This would hamper full disclosure and limit the ability to compare the data from multiple organizations. Using data from government organizations could overcome this difficulty but that would exclude a very large segment of the information system world.

Experiment Guidance

The workshop participants developed several suggestions to guide experiments exploring information system phenomena. Although the participants generated some general advice, IA represented the primary context for this guidance.

- Exploit use-cases for IS data
- Examine different IS conditions
- Examine IS behavior near equilibrium points
- Choose experimental facilities to answer the question

- Use different experiments to characterize different phenomenological mappings
- Conduct different experiments for different attacks
- Choose between simulation and emulation
- Use successively refined experiments
- Check experimental results against real system observations
- Define meaningful levels of phenomenological aggregation
- Begin with strongly stated hypotheses

Explore Use-Cases for IS Data. Dr. Benzinger suggested exploring use-cases to as a source of data about information flows through a system. Although use-cases are less formal than good experiments and the data they provide may be more difficult to rigorously interpret, they are a commonly accepted method within the software development community as a data source and do provide insight into aspects of information system behavior that may require significantly more effort to obtain through formal experiments.

Examine Different IS Conditions. Dr. Benzinger noted that many system vulnerabilities occur only when an information system encounters stressful conditions such as maintenance activities and heavy loads. These conditions create opportunities for attackers and must be considered in any experiments performed with the goal of providing useful information for IA. She suggested that initial experiments could begin in a scripted mode to push the system into the stress condition. Once in this condition, the experiments could explore the effects of different types of attacks. Additional experiments could explore the conditions needed to precipitate cascading failures and assess how the system vulnerability changes.

Examine IS Behavior near Equilibrium Points. Dr. Belyavin suggested that experiments should define how an information system behaves near its equilibrium points (i.e., those points where it is stable). This provides the basic information needed to identify where a set of attractors is located and how the system behaves as it moves from one attractor to the next. This approach is especially valuable when the system exhibits chaotic behavior and it enables modelers to apply the mathematics of chaotic systems in their analyses.

Choose Experiment Facilities to Answer the Question. The workshop participants directed considerable discussion toward the characteristics of experimental facilities. Not surprisingly, they concluded that the range of experiment opportunities and the lack of a broad base of experience in conducting controlled experiments upon information systems makes any guidance for selecting experimental facilities extremely sensitive to the goals of the experiments. In some cases, experiments may require only limited capabilities. Others may require the ability to control specific aspects of information system conditions. Many phenomena behind complex information system behavior may make observation of the behavior of actual complex

information systems the only acceptable choice. Dr. Belyavin observed that where one wants to run experiments depends entirely upon what part of the system they are considering.

Use Different Experiments to Characterize Different Phenomenological Mappings. Dr. Gopal observed that different mappings may exist between observable system properties and the macroscopic phenomena governing an information system's behavior. A single experiment will likely not provide sufficient insight to untangle all of the possible mappings. Some phenomena may require several different experiments, each that explore a different mapping or set of mappings, to adequately characterize those phenomena. In some cases, experimenters should attempt to understand system behavior at a small scale then extend those results to larger scales of implementation. This observation strengthens the argument for a carefully planned experimental program, consisting of several individual but interdependent experiments in order to obtain clearly interpretable results about information system phenomena.

Conduct Different Experiments for Different Attacks. Dr. Belyavin extended the different experiments for different mappings to the IA problem. He suggested that several different classes of attacks and attackers exist and that these may require different experimental means to understand the effectiveness of protection against those attacks. This argument emphasizes the need to clearly identify the purpose of the experiment(s) during the planning process. Indefinitely defined experimental goals will probably lead to experiments that produce data that is difficult to interpret. Experiment planning goes far beyond conceiving a testable hypothesis. It must also identify the criteria and the methods used to perform the tests as well as the mechanisms that could lead to incorrect answers (i.e., error sources). Different attacks may exploit totally different information system phenomena. Further, different protective mechanisms to the same attack may capitalize upon different phenomena still. The interactions between these two perspectives lead to a matrix of possible interactions, each of which may represent a different set of phenomena. Any experiments seeking to characterize these interactions must tease the influences of the different, perhaps interdependent, phenomena apart.

Choose between Simulation and Emulation. In some cases, performing experiments in simulation may provide the most cost effective means to characterize particular phenomena, especially if the simulation's level of abstraction and fidelity matches the experimental goals suitably. In other cases, the true complexity of a real system may only be available through emulation, i.e., using another real information system to represent the information system of interest. For example, one could use the DARPA TIC to emulate parts of the Global Command and Control System (GCCS). An experimenter must consider this tradeoff carefully. Simulations are easy to control and observe but their efficacy depends upon their validity. Emulations are much easier to control and observe than real systems but they, like simulations, represent abstractions of real situations and must be used carefully. Further, emulations can involve much greater complexity than simulations and may introduce stochastic behavior that must be characterized. Emulations can represent situations much closer to those experienced in real systems since they can involve the

actual equipment and people. But, their added complexity can make their data difficult to interpret.

Use Successively Refined Experiments. One solution to dealing with experiment complexity and the difficulty associated with interpreting the results from complex experiments came from Mr. Gross when he suggested planning experiments that successively refined the search space of the experiment. This approach works in both directions. One could conduct a large scale experiment involving may information system phenomena to assess the magnitude of the aggregated effects of those phenomena and then successively narrow the experiment scope to tease the effects of each phenomena apart. One could also perform narrow experiments upon carefully partitioned segments of the system behavior space then perform successively broader experiments to assess the magnitude of the interactions between phenomena.

Check Experiment Results against Real System Observations. Dr. Belyavin defined one important relationship between controlled experiments, simulations and observations of the actual system behavior. He suggested that any results from controlled experiments and simulations must be checked against the behavior of the real system to ensure their validity. In other words, one purpose for observations of the real system behavior is to gauge the accuracy of the experimental or simulation results. In some cases (e.g., system recovery), the actual system behavior may be the only source of accurate information due to the inherent behavioral complexity.

Define Meaningful Levels of Phenomenological Aggregation. Prof. Toffoli asserted that part of any experimental program should be aimed at understanding where the natural levels of aggregation exist for the phenomena and applications being examined (e.g., IA). He described that all physical systems seem to have distinct levels of aggregation that are meaningful to the phenomena being considered (e.g., atoms for chemistry, nuclear components for nuclear physics, stars and planets for astronomy, and galaxies for cosmology). IA experimenters must find the right levels of aggregation of information system phenomena for their problems. Finding these will simplify the experiments and strengthen the validity and applicability of their results. Prof. Clarke suggested that specific experiments may be needed to find these aggregation levels for information systems. These experiments should define the limits of validity of experimental results at those levels.

Begin with Strongly Stated Hypotheses. Dr. Belyavin warned of the need to have clearly and strongly stated hypotheses about information system phenomena in order to correctly identify experiment requirements. Broadly or weakly stated hypotheses will lead only to experiments that confirm prior prejudices and from which very little, if anything, will be learned. While boldly stated hypotheses risk the possibility of experimental refutation, they clearly identify the model they propose to characterize information an system phenomenon. Unambiguous refutation of this explanation narrows the space to search for better models of information system phenomena. Further, well planned experiments can also suggest modifications to the hypothesized model that could lead to satisfactory explanations.

Validation of IA-Related Theories

"Lots of models exist but very little data exists to support the correctness of those models." Mr. David Gross

This comment fairly captures the current state of the science of information systems. Dr. Benzinger went on to say "Artifacts generally precede theory. We have examples of secure components but no general theory to describe them. In many cases, we might possibly derive a general theory from the existing artifacts."

The next to the last working session encouraged the participants to discuss how they might approach validating their own theories describing the macroscopic phenomena of complex information systems. Table 10, below, summarizes the results of this discussion.

Table 10. Possible Approaches to Validate Individual Participant Theories.

Participant Theory	Possible Experimental Validation Approach
Kolmogorov Complexity	use Swarm simulations to represent a closed information system; measure the Kolmogorov complexity of that system and observe how that complexity evolves over time under different conditions; check the validity of the resulting model against observations of the Internet
EPI	use observations of Internet and, possibly, Swarm simulations to test EPI predictions of oscillatory behavior in attacker-defender interactions (i.e., the information gain problem)
Cyberlogic	use existing data on the failures and compromises of real systems to validate the cyberlogic theory; apply that theory to the design of real systems to determine if it improves the design process or products
IAM	may use experiments to test the definitions for information system composition operators but IAM does not lead naturally to experiments because its models are derived from rigorous abstractions and only depend upon the assumptions underlying those abstractions
Composite Systems	build models of team behavior from existing data on team performance then use simulation and controlled experiments to test those models further

Table 10. Possible Approaches to Validate Individual Participant Theories (continued).

Participant Theory	Possible Experimental Validation Approach
Category Theory	use existing data from networked simulation experiments and information system attack history to test the validity of the commutivity criterion for interoperability
Cyberspace Cartography	use existing data from past Red Team experiments (e.g., RT-001) to test predicted adversarial course of action probabilities; some of this testing has been done and looks pretty good; also, use controlled experiments to explore cost-benefit analysis
Information Physics	use controlled experiments to develop a consistent definition of information work across different implementations (e.g., electronic computers, people); use existing data on information system behavior and controlled experiments to test the hypotheses that information flow rate is proportional to goal force magnitude and diffusion constants; use existing data on information system behavior and controlled experiments to test the hypothesis that information work is proportional to information complexity change

During this discussion, some of the participants responded to their own or others' validation approaches.

- Dr. Cox sees two opportunities for experiments exploring EPI. One of these would determine whether EPI applies to describing information system behavior. The other determines how well EPI applies specifically to such IA problems as information gain between attackers and defenders. Both of these experiment opportunities would test the accuracy of EPI predictions but in different contexts, one broadly applicable to all information systems and the other applicable specifically to IA.
- Dr. Cox also noted that a relationship exists between Fisher information and Kolmogorov complexity, and that it may be as strong as formal equivalence. As a result, the experimental work on Kolmogorov complexity should have direct relevance to the EPI work and vice versa.
- Regarding the experiments considering Kolmogorov complexity, Dr. Belyavin suggested that "We must look at least three points across a range to determine if the Kolmogorov complexity is the right sort of metric. You might need to drive the system quite hard to observe a broad enough range."

- Dr. Bush asked how the suggested Cyberlogic experiments would quantify their value.
- Dr. Benzinger intends to take the IAM theory directly to technology transfer. She will develop the equations to guide design decisions and provide those equations for those people who want to use them.
- Mr. Harmon recommended that experiments could help to validate the IAM equations and determine the limits of the validity of the underlying assumptions.

UML as Theory Notation

The workshop participants discussed using the Unified Modeling Language (UML) as a means to describe information system behavior and phenomena. The UML notation is very powerful and is becoming widely accepted in the information systems community. As a result, it may provide a good means to capture and communicate the essence of information system theories. In this context, Mr. Frentz was curious about whether other people were using UML and what tool environment they had to support that use. He said the BBN was currently using UML with the Rational toolset. Dr. Cox asked people at Sandia and found little expressed interest in UML. Mr. Gross said that parts of Boeing were using UML and that the Rational toolset was the Boeing standard. Dr. Gopal confirmed that that was also true for Honeywell. This limited survey suggests that UML together with the Rational toolset may provide a useful mechanism to express information theory. It may also support different aspects experiment planning (e.g., experiment arrangement description, data analysis technique planning and description).

Conclusions and Recommendations

The final working session focussed upon developing a set of conclusions and recommendations upon which all of the participants agreed. The discussion below presents the results of this session.

One of the original goals of this workshop was to recommend a prioritized list of experiments that would best resolve existing theoretical conflicts, identify the most promising theory, and advance theory into unknown areas. After considerable discussion, the workshop participants agreed that this goal was premature given the state of current theory.

Workshop Conclusions

The workshop participants agreed unanimously that better knowledge of the scientific fundamentals of information systems will improve the state of IA practice in the following ways.

- Identifying meaningful measurements through which to detect and characterize attacks and failures
- Determining the causal relationships that would enable inferring information system state from measurements
- Enabling effective assessments of information system performance for both functional and support activities

The participants felt that sufficient theory describing information system phenomena exists in narrow areas that it could be further explored with small focused experiments. These experiments would

- Lead to refinements of theory and
- Help to resolve applied work.

The participants cautioned that the science of information systems is indeed in its infancy. The lack of a common vocabulary and the uneven characterization of the mappings of microscopic variables to macroscopic phenomena indicate this immaturity. However, substantial theoretical affinities exist with well developed disciplines (e.g., mathematics, physics, economics, sociology, evolution). These affinities could be leveraged to rapidly accelerate the maturing of the science describing information system behavior. On the other hand, despite these affinities, none of these well establish disciplines are enough without further development. For example,

- General systems theory does not address the issue of information content;
- Information and network theory do not address non-equilibrium conditions; and
- Intimate integration of the human element with information systems presents additional challenges.

Workshop Recommendations

Building upon these conclusions, the workshop participants formulated several recommendations for the information system community in general and DARPA specifically.

- **Aggressively Pursue the Science Explaining Information System Phenomena.** A coherent and consistent body of knowledge, scientific knowledge, describing the macroscopic phenomena of complex information systems must be developed, either now or later. The absence of such knowledge will ultimately limit our ability to create and maintain information systems with predictable and assured levels of functionality and performance. The commercial world will likely not develop this technology base independent of Government funding because its return is too long term and because few commercial organizations support research arms to perform such basic studies. These considerations lead to the recommendation that the Government should maintain a meaningful set of projects focussed upon developing the science relevant to IA. While this recommendation is no surprise coming from a workshop focussed upon the science of information systems, that lack of surprise in no way diminishes its importance and does not change the fact that the lack of this scientific knowledge will impede, and ultimately prevent, the predictable implementation of reliable complex information systems.
- **Emphasize Experimentation.** A core element of any scientific program must emphasize experimentation to validate the results of the theoretical component, to deconflict the descriptions of competing theories, and to provide data characterizing the phenomena of interest. Developing a science describing information systems is no exception to this general guidance. This widely appreciated wisdom leads to the recommendation that the Government should support the experimentation necessary to promote development of valid, comprehensive and relevant theory. Some theory already exists and supporting the experimentation to test this theory can add it to the base of knowledge useful in developing practical information systems.
- **Define Consistent Terminology.** The participants toiled at making sure they understood one another and felt that this problem pervades the entire field of information systems. The lack of a common and consistent vocabulary to describe information system characteristics and phenomena will hinder both scientific and engineering development of the field. Practitioners must get their terms crisply defined so they can meaningfully compare and contrast their theories and so they can determine the types of experiments needed to test these theories.
- **Build upon Existing Knowledge.** The recommended scientific endeavor must build upon the substantial body of relevant theory that exists from other well developed disciplines. This recommendation, while sensible, seems obvious but the most important point is that an entire scientific discipline does not need to be built from scratch to realize a practical understanding of information system behavior. Considerable knowledge, both theoretical and experimental, exists that provides a firm foundation for the science

characterizing information system phenomena. Failing to exploit every shred of this existing knowledge base would be wasteful and may lead to inconsistencies with other scientifically robust areas.

- **Identify and Fill Theory Holes.** The science of information systems does not lack theory. Some comes from existing related disciplines such as mathematics and systems theory. Some comes from recent efforts to explain different aspects of information system phenomena. But, the union of this body of theory still only sparsely addresses the areas of practical interest. The recommended program should define the limits of the existing theoretical knowledge as well as those areas where theories conflict. These limits form the boundaries upon which further knowledge can be added. The space between isolated closed boundaries and those areas where theories with overlapping boundaries conflict define where experimentation is necessary.
- **Link Scientific Discovery to Practical Development.** These recommendations do not argue for support of science purely for the sake of altruistically adding to humanity's base of knowledge. The participants felt that, like engineering programs, scientific programs can be productive while contributing to specific applications. Assuming IA represents the practical domain of interest, identifying where current theory can bear on solving IA problems, specifically the challenge problems discussed below, will further draw the theoretical development toward practical applications. This will help to focus the recommended program and improve its perceived productivity. Prof. Toffoli characterized the body of existing theory as three animals: one representing information systems operating at equilibrium, one representing information systems operating away from equilibrium in a dissipative but steady state, and the final one representing deliberate design that starts from a rationale and produces a useful product. Any scientific program should hitch those three animals to a practical application, such as IA. In this way, the recommended program should nurture development of those fundamentals that enable requisite improvements in building the critical systems of the future.
- **Prioritize IA Problems.** Again, assuming that IA represents the practical domain of interest, we must prioritize what is important about IA. This component will lead to a list of IA challenge problems, or even a grand challenge problem, that will focus the recommended research program and improve the likelihood of immediately useful results from it. The unknowns of this scientific domain are so broad that any effort, even an application-focussed one, will contribute to scientific knowledge. Clearly defined challenge problems can act as the beacons to aid unfaltering navigation past the temptations to wallow in pure science and lead to science that makes a significant difference in existing practice.

REFERENCES

[1] Marion Severynse et al., eds., Webster's II, New College Dictionary, Houghton Mifflin Co., Boston, MA, 1995.

[2] Jane Radatz et al., eds., The IEEE Standard Dictionary of Electrical and Electronics Terms, 6th Edition, IEEE Std 100-1996, IEEE Press, New York, NY, 8 April 1997.

[3] Sybil P. Parker et al., eds., McGraw-Hill Dictionary of Engineering, McGraw-Hill Publishing, New York, NY, 1997.

[4] Teri Kieffer et al., eds., Microsoft Press Computer User's Dictionary, Microsoft Press, Redmond, WA, 1998.

[5] R.W. Keyes & Rolf Landauer, "Minimal Energy Dissipation in Logic," IBM Journal of Research and Development, 14, 1970, pp152-157.

[6] S. Y. Harmon, Exploring a Theory Describing the Physics of Information Systems, Final Report, Zetetix, Oak Park, CA, November 2000.

[7] Gregg Schudel, DARPA Information Assurance Laboratory Facilities and Capabilities Document, Revision 1.0, BBN Technologies/GTE, Arlington, VA, 12 June 2000.

ACKNOWLEDGEMENTS

This workshop would not have been possible without the support of the Air Force Research Laboratories, Rome NY and the Defense Advanced Research Projects Agency and its Information Systems Office. The author greatly appreciates the support, guidance and encouragement of Mr. Michael Skroch, Manager for the Information Assurance Science and Engineering Technology Program and Ms Deborah Cerino of the Air Force Research Laboratories. Further, this workshop would not have been possible without the diverse assistance of Mr. Sam Nicholson, Schafer Corporation, and Dr. David Tseng.

*U.S. GOVERNMENT PRINTING OFFICE: 2001-610-055-10091

MISSION
OF
AFRL/INFORMATION DIRECTORATE (IF)

The advancement and application of Information Systems Science and Technology to meet Air Force unique requirements for Information Dominance and its transition to aerospace systems to meet Air Force needs.